Catalytic subunit of human telomerase and its diagnostic and therapeutic use

Structure and function of the chromosome ends

5   The genetic material of eukaryotic cells is distributed on Linear chromosomes. The ends of these hereditary units are termed telomeres, derived from the Greek words *telos* (end) and *meros* (part or segment). Most telomeres consist of repeats of short sequences which are mainly constructed from thymine and guanine (Zakian, 1995). The telomere sequences of related organisms are often similar and these sequences are even conserved between species which are more phylogenetically remote. It is a remarkable fact that the telomeres are

10   constructed from the sequence TTAGGG in all the vertebrates which have so far been examined (Meyne *et al.*, 1989).

The telomeres exert a variety of important functions. They prevent the fusion of chromosomes (McClintock, 1941) and consequently the formation of dicentric hereditary

15   units. Chromosomes of this nature, possessing two centromeres, can lead to the development of cancer due to loss of heterozygosity or the duplication or loss of genes.

In addition, telomeres serve the purpose of distinguishing intact hereditary units from damaged hereditary units. Thus, yeast cells ceased dividing when they harboured a

20   chromosome which lacked a telomere (Sandell and Zakian, 1993).

Telomeres carry out another important task in association with DNA replication in eukaryotic cells. In contrast to the circular genomes of prokaryotes, the Linear chromosomes of eukaryotes cannot be completely replicated by the DNA polymerase complex. RNA primers

25   are required for initiating DNA replication. After the RNA primers have been eliminated and the Okazaki fragments have been extended and then ligated, the newly synthesized DNA strand lacks the 5' end because the RNA primer at that point cannot be replaced with DNA. For this reason, without special protective mechanisms, the chromosomes would shrink with

30   every cell division ("end-replication problem", Harley *et al.*, 1990). The non-coding telomere

sequences probably represent a buffer zone for preventing the loss of genes (Sandell and Zakian, 1993).

Over and above this, telomeres also play an important role in regulating cell ageing (Olovnikov, 1973). Human somatic cells exhibit a limited capacity to replicate in culture; after a certain time they become senescent. In this condition, the cells no longer divide even after being stimulated with growth factors; however, they do not die but remain metabolically active (Goldstein, 1990). Various observations provide support for the hypothesis that a cell determines from the length of its telomeres how often it can still divide (Allsopp et al., 1992).

In summary, the telomeres consequently possess central functions in the ageing of cells and in the stabilization of the genetic material and prevention of cancer.

The enzyme telomerase synthesizes the telomeres

As described above, organisms possessing Linear chromosomes are only able to replicate their genomes incompletely in the absence of a special protective mechanism. Most eukaryotes use a special enzyme, i.e. telomerase, to regenerate the telomere sequences. Telomerase is expressed constitutively in the single-cell organisms which have so far been examined. By contrast, in humans, telomerase activity was only detected in germ cells and tumour cells whereas neighbouring somatic tissue did not contain any telomerase (Kim et al., 1994).

Telomerase in ciliates

Like the telomeres, telomerase was identified for the first time in the ciliate *Tetrahymena thermophila*. Telomerase activity was detected by extending the single-stranded oligonucleotide d(TTGGGG)$_4$ in the presence of dTTP and dGTP (Greider and Blackburn, 1985). In this reaction, the *Tetrahymena* telomere sequence TTGGGG was added repeatedly to the primer. Even when an oligonucleotide having the irregular telomere sequence of *Saccharomyces cerevisiae*, T(G)$_{1-3}$, was offered as the starting material, the telomerase

extended the primer with the telomere sequence of Tetrahymena (Greider and Blackburn, 1985). From these results, it was concluded that the telomerase itself carries the template for the sequence of the telomeres.

5   Once the existence of an RNA component in the telomerase had initially been demonstrated (Greider and Blackburn, 1987), the gene for the RNA subunit of the telomerase was cloned a short while later (Greider and Blackburn, 1989). This RNA contains a region which is complementary to the Tetrahymena telomere sequence (termed "complementary region" below). The activity of the telomerase depended on the RNA component, as was demonstrated by digesting the RNA, leading in turn to subsequent loss of activity. If the

10   complementary region of the telomerase RNA was mutated, the corresponding mutations were incorporated in vivo into the Tetrahymena telomeres (Yu et al., 1990). Telomerase consequently belongs to the class of RNA-dependent DNA polymerases.

15   The first protein subunits of the Tetrahymena telomerase, i.e. p80 and p95, were identified in 1995 (Collins et al., 1995). The observation that p95 anchors the enzyme to the DNA and p80 binds the RNA component led to the following model: the telomerase RNA anneals by its complementary region to the single-stranded 3' overhang. The 3' overhang is extended by incorporating the corresponding nucleotides in the 5'-3' direction. The de novo synthesis of telomeres probably involves an elongation step and a translocation step. Once a telomere

20   sequence has been synthesized, the telomerase presumably moves along the DNA until it is once again in a position to be able to add a complete telomere sequence. This model does not have to be generally valid since great differences exist between the telomerases of different species with regard to the number of nucleotides which the enzyme adds before it dissociates

25   from the telomere (Prowse et al., 1993).

In addition to this, telomerase subunits from other organisms have also recently been identified. Two protein subunits, i.e. p123 and p43, which do not exhibit any homology with the Tetrahymena telomerase proteins, have been found in the ciliate Euplotes aediculatus.

30   The telomerase subunit p123 exhibits a basic domain at its N terminus and a domain for a reverse transcriptase (RT) at the C terminus, suggesting this protein has a catalytic function,

(Lingner *et al.*, 1997). Furthermore, p123 has been reported to share significant homology with the *Saccharomyces cerevisiae* protein Est2 which was found by Lundblad (Lingner *et al.*, 1997).

Whereas p80 and p95 have not hitherto been demonstrated to possess any function which is essential for telomerase activity, the potential catalytic telomerase subunits p123/est2p have been unambiguously shown to have a key function: mutation of the active centre of the est2p RT led to significant truncation of the telomeres in yeast cells (Lingner *et al.*, 1997).

Telomerase components from mammalian cells

The RNA components of the telomerases of various organisms, inter alia of *Saccharomyces cerevisiae*, mice and humans (Singer and Gottschling, 1994; Blasco *et al.*, 1996; Feng *et al.*, 1995), have by now been cloned. All the telomerase RNAs known to date comprise a region which is complementary to the telomere sequence of a particular organism. However, the primary sequence of the human telomerase RNA (hTR) does not display any similarity to the RNA components of the ciliates or of *Saccharomyces cerevisiae*. On the other hand, regions exist which are conserved between human and murine telomerase RNA (Feng *et al.*, 1995).

The isolation of a human telomerase-associated protein (hTP1) has recently been described (Harrington *et al.*, 1997). On the basis of its homology with the *Tetrahymena* telomerase p80 subunit, the corresponding gene was found in an EST data base which is not available to the general public (Harrington *et al.*, 1997). hTP1 is composed of 2627 amino acids and, in the N-terminus, exhibits three domains which possess at most 46% homology with p80. 16 repeats of the amino acids tryptophan and asparagine, which presumably mediate a protein/protein interaction, were shown to be present, as an additional structural element, in the C-terminal region.

Activation of the telomerase in human tumours

In humans, it was originally only possible to demonstrate telomerase activity in germ line cells and not in normal somatic cells (Hastie *et al.*, 1990; Kim *et al.*, 1994). After a more sensitive detection method had been developed (Kim *et al.*, 1994) a low level of telomerase activity was also detected in hematopoietic cells (Broccoli *et al.*, 1995; Counter *et al.*, 1995;

5   Hiyama *et al.*, 1995). However, these cells nevertheless exhibited a reduction in the telomeres (Vaziri *et al.*, 1994; Counter *et al.*, 1995). It has still not been clarified whether the quantity of enzyme in these cells is insufficient to compensate for the telomere loss or whether the measured telomerase activity stems from a subpopulation, e.g. of incompletely differentiated $CD34^+38^+$ precursor cells (Hiyama *et al.*, 1995). In order to clarify this point, it would be

10   necessary to detect the telomerase activity which was present in a single cell.

Interestingly enough, however, significant telomerase activity has been detected in a large number of the tumour tissues which have been tested to date (1734/2031, 85%; Shay, 1997), whereas no activity has been found in normal somatic tissue (1/196, <1%, Shay, 1997). In

15   addition, a variety of investigations demonstrated that the telomeres continued to shrink in senescent cells which were transformed with viral oncoproteins and that it was only possible to find telomerase in the subpopulation which survived the growth crisis (Counter *et al.*, 1992). The telomeres were also stable in these immortalized cells (Counter *et al.*, 1992). Similar findings derived from investigations in mice (Blasco *et al.*, 1996) support the

20   assumption that reactivation of the telomerase is a late event in tumorigenesis.

Based on these results, a "telomerase hypothesis" was developed which links the loss of telomere sequences and cell ageing to telomerase activity and the genesis of cancer. In long-lived species such as humans, the shrinking of the telomeres can be regarded as a tumour

25   suppression mechanism. Differentiated cells, which do not contain any telomerase, cease dividing when the telomeres have reached a particular length. If such a cell mutates, a tumour can only develop from it if the cell is able to extend its telomeres. Otherwise, the cell would continue to lose telomere sequences until its chromosomes became unstable and it finally died. Reactivation of the telomerase is presumably the main mechanism which tumour cells

30   deploy in order to stabilize their telomeres.

It follows from these observations and ideas that it should be possible to develop a therapy for tumours based on inhibiting telomerase activity. Conventional cancer therapies using cytostatic agents or short-wave irradiation damage all the dividing cells in the body in addition to damaging the tumour cells. However, since it is only germ line cells which contain significant telomerase activity, apart from tumour cells, telomerase inhibitors would attack the tumour cells more specifically and consequently evoke fewer undesirable side effects. Since telomerase activity has been detected in all the tumour tissues tested to date, it would be possible to employ these therapeutic agents against all types of cancer. The effect of telomerase inhibitors would then set in when the telomers of the cells had shortened to such an extent that the genome had become unstable. Since tumour cells usually exhibit shorter telomeres than do normal somatic cells, it would be cancer cells which would first of all be eliminated by telomerase inhibitors. By contrast, cells possessing long telomeres, such as the germ cells, would not be damaged until a much later stage. Telomerase inhibitors consequently represent an approach which points the way forward for cancer therapy.

However, it will only be possible to provide unambiguous answers to questions regarding the nature and the points of attack of physiological telomerase inhibitors when the protein structures of the enzyme, together with their functions, have also been identified and a deeper understanding of the various telomere-binding proteins has been obtained.

The invention relates to the catalytically active human telomerase subunit (phTC), where appropriate in purified form, to active moieties of the protein, to modulators, in particular agonists of the protein, to substances which imitate the function of the protein and to combinations of these components.

The invention furthermore relates to:

- The nucleic acid sequence which encodes the human protein phTC, specifically:

  - the genomic sequence of the hTC gene,
  - the cDNA sequence of the hTC gene,

- the DNA sequence of hTC variants

- the sequence of the mRNA which is transcribed from the hTC gene,

- parts of the abovementioned sequences, including the DNA sequence (SEQ ID No. 1) of hTC which is shown in Fig. 1.

5

- The nucleic acid sequences which encode hTC-homologous proteins in other mammals, specifically:

- the genomic sequences of hTC-homologous genes,

- the cDNA sequences of hTC-homologous genes,

10

- the sequences of the mRNAs which are transcribed from hTC-homologous genes,

- parts of the abovementioned sequences.

15 - Nucleic acid sequences which, in humans and other mammals, encode proteins which are related to the phTC protein, specifically:

- the genomic sequences of hTC-related genes in humans and other mammals,

- the cDNA sequences of hTC-related genes in humans and other mammals,

- the sequences of the mRNAs which are transcribed from hTC-related genes in humans and other mammals,

20

- parts of the abovementioned sequences.

- The above-described phTC protein, which is isolated from mammalian cells (cf. Fig. 2 and SEQ ID No. 2).

25

- The phTC protein which is labelled with a detection reagent, with the detection reagent preferably being an enzyme, a radioactively labelled element or a fluorescent chemical.

30

- An antibody which is directed against the phTC protein.

According to a preferred embodiment, this antibody is a polyclonal antibody.

According to another preferred embodiment, this antibody is a monoclonal antibody.

Antibodies of this nature can be produced, for example, by injecting a host, which is substantially immunocompetent, with a quantity of a phTC polypeptide, or a fragment thereof, which is effective for producing the antibody, and by subsequently isolating this antibody.

In addition, an immortalized cell line which produces monoclonal antibodies can be obtained in a manner known per se.

Where appropriate, the antibodies can be labelled with a detection reagent.

Fragments which possess the desired specific binding properties can also be employed instead of the complete antibody.

Preferred examples of such a detection reagent are enzymes, radioactively labelled elements, fluorescent chemicals or biotin.

Oligonucleotides in purified form which have a sequence which is identical or exactly complementary to a contiguous sequence, of from 10 to 500 nucleotides in length, of the above-described genomic DNA, cDNA or mRNA.

An oligonucleotide of this nature can, in particular, be an oligodeoxy-ribonucleotide or an oligoribonucleotide or a peptide nucleic acid (PNA).

Preference is given to oligonucleotides which inhibit, repress or block the activity of the telomerase when they bind to the hTC mRNA.

A DNA sequence, or a degenerate variation of this sequence, which encodes the phTC protein, or a fragment of this protein, where appropriate comprising the DNA sequence in Figure 1a, or a DNA sequence which hybridizes with the previously cited DNA sequence under standard hybridization conditions.

A recombinant DNA molecule which comprises a DNA sequence, or a degenerate variation of this sequence, which encodes phTC or a fragment of phTC, with the latter sequence preferably comprising the DNA sequence in Figure 1a, or which comprises a DNA sequence which hybridizes with the previously cited DNA sequence under standard hybridization conditions.

In the abovementioned recombinant DNA molecule, the described DNA is preferably linked to an expression control sequence.

Examples of expression control sequences which are particularly preferred are the early or late promoter of the SV40 virus or adenovirus, the lac system, the trp system, the TAC system, the TRC system, the main operator and promoter regions of phage λ, the control regions of the fd coat protein, the 3-phosphoglycerate kinase promoter, the acid phosphatase promoter and the yeast α-mating factor promoter.

A single-cell host which has been transformed with the above-described recombinant DNA molecule which comprises the DNA sequence, or a degenerate variation of this sequence, which encodes the phTC protein or a part of this protein. In this recombinant DNA molecule, the said DNA sequence is linked to an expression control sequence.

Preferred examples of the single-cell host are: E. coli, Pseudomonas, Bacillus, Streptomyces, yeasts, CHO, R1.1, B-W, L-M, COS 1, COS 7, BSC1, BSC40 and BMT10 cells, plant cells, insect cells and mammalian cells in cell culture.

A recombinant virus which is transformed with one of the previously described DNA molecules or a derivative or fragment of this molecule.

A method for inhibiting telomerase activity in human cells, preferably neoplastic cells, in which an exogenous polynucleotide which consists of a transcription unit is transferred into the cells. This transcription unit comprises a polynucleotide sequence of at least 29 consecutive nucleotides, which sequence is substantially identical or substantially complementary to the hTC RNA sequence and is linked to a heterologous transcription-regulating sequence which controls the transcription of the linked polynucleotide in the said cells.

Preferably, the abovementioned heterologous transcription-regulating sequence comprises a promoter which is constitutively active in human cells.

Alternatively, the heterologous transcription-regulating sequence can comprise a promoter which can be induced or repressed in human cells by adding a regulatory substance. Examples of such promoters are inducible and repressible tetracycline-dependent promoters, heat shock promoters and metal ion-dependent promoters.

The abovementioned exogenous polynucleotide can, for example, be a viral genome containing a transcription unit from the human hTC DNA component.

Particularly preferably, the said transcription unit produces antisense RNA which is substantially complementary to the human hTC RNA component.

Particular preference is also given to the exogenous polynucleotide being able to comprise the sequence in Fig. 1a.

A polynucleotide for the genetic therapy of a human disease. This polynucleotide consists of a transcription unit which comprises a polynucleotide sequence of at least 9 consecutive nucleotides, which sequence is substantially identical or substantially

complementary to the hTC RNA sequence and is linked to a heterologous transcription-regulating sequence which controls the transcription of the linked polynucleotide in said cells.

5   -   A method for detecting telomerase-associated conditions in a patient, which method comprises the following steps:

A. Detecting the phTC protein in body fluids or cell samples in order to obtain a diagnostic value;

10   B. Comparing the diagnostic value with standard values for the phTC protein in standardized normal cells or body fluids of the same type as the test sample;

C. Detecting diagnostic values which are higher or lower than the standard comparative values and which indicate a telomerase-associated condition, which condition in turn indicates a pathogenic condition.

15   This method is preferably employed for detecting a neoplastic disease in a patient. The method then comprises the following steps:

A. Detecting the phTC protein in cell samples in order to obtain a diagnostic value;

B. Comparing the diagnostic value with standard values for the phTC protein in

20   non-neoplastic cells of the same type as the test sample;

C. Diagnostic values which are clearly higher than standard comparative values indicate a neoplastic condition.

25   -   A method for determining the presence of the phTC protein in a cell or cell sample, which method is based on amplifying an hTC polynucleotide, or hybridizing an hTC polynucleotide, a primer or an hTC-complementary sequence with an hTC polynucleotide.

30   -   A test kit for detecting phTC in cell samples and body fluids, with it being possible, for example, for labelled, immunochemically-reactive components to be: polyclonal

antibodies against phTC, monoclonal antibodies against phTC, fragments of these antibodies or a mixture of these components.

A method for preventing and/or treating cell disturbance or destruction and/or malfunction and/or other symptoms in humans, which method is based on administering a therapeutically effective quantity of catalytically active human telomerase, its functional equivalents or its catalytically active fragments. It is also possible to conceive of using a substance which promotes the production and/or activity of phTC; a substance which can imitate the activity of phTC; a substance which can inhibit the production and/or activity of phTC, or a mixture of these substances. A specific binding partner can also be employed.

The method is preferably employed for preventing or treating ageing or cancer diseases.

Substances which are able to affect the activity of phTC, i.e. inhibit or promote, are here termed modulators. Such modulators can be found, in a manner known per se, by testing their effect on telomerase activity in a telomerase assay. Examples of telomerase assays are given in Example 15.

Modulators of phTC are of interest for treating diseases which are connected with telomerase. The prevention or treatment of ageing processes or of cancer diseases may, in particular, be mentioned in this context.

An antisense nucleic acid against the hTC mRNA, which nucleic acid comprises a nucleotide sequence which hybridizes with the said mRNA, with the antisense nucleic acid being an RNA or a DNA.

Preferably, the antisense nucleic acid binds to the start codon of the particular mRNA.

A recombinant DNA molecule which contains a DNA sequence from which an antisense ribonucleic acid against the hTC mRNA is produced during transcription. This said antisense ribonucleic acid comprises a nucleic acid sequence which can hybridize to the said hTC mRNA.

A DNA molecule of this nature can be used to prepare a cell line having a reduced expression of phTC by transfecting a phTC-producing cell line with this recombinant DNA molecule.

A ribozyme which cleaves the hTC mRNA.

This ribozyme is preferably a *Tetrahymena*-type ribozyme or a hammerhead-type ribozyme.

A recombinant DNA molecule which contains a DNA sequence whose transcription leads to the production of a ribozyme of this nature.

This recombinant DNA molecule can be used to transfect a phTC-producing cell line.

A combination which consists of a pair of human hTC polynucleotide PCR primers, with the primers preferably consisting of sequences which correspond to the sequence of the human hTC mRNA or which are complementary to this sequence.

A combination which comprises a polynucleotide hybridization probe for the human hTC gene, with the probe preferably comprising at least 29 consecutive nucleotides which correspond to the sequence of the human hTC gene or which are complementary to this sequence.

Animal models which can be used to investigate telomerase/telomere regulation *in vivo*. Thus, tumour development and ageing can, for example, be directly investigated using knockout animals or transgenic animals.

In the case of proteins or peptides, functional equivalents are those compounds which, while being distinguishable with regard to amino acid sequence, essentially have the same functions.

5 Known examples of these compounds are isoenzymes or so-called microheterogeneities in proteins.

In the case of the oligonucleic or polynucleic acids, functional equivalents are to be understood as being those compounds which differ in nucleotide sequence but which encode 10 the same protein. The existence of such compounds may be attributed, for example, to the fact that the genetic code is degenerate.

Explanation of the figures:

15 Fig. 1:   cDNA sequence of the catalytic subunit of human telomerase (hTC) (SEQ ID No. 1).

Fig. 2:   Amino acid sequence which is deduced from the hTC DNA sequence depicted in Fig. 1 (SEQ ID No. 2).

20 The DNA sequence depicted in Fig. 1 can be completely translated from Position 64 to Position 3461 into an amino acid sequence. The amino acid residues are depicted in accordance with their single-letter code.

25 Fig. 3:   Ethidium bromide-stained agarose gel containing AA281296 DNA which has been treated in different ways.

The figure shows an ethidium bromide-stained 0.8% agarose gel. Two different DNA size standards are loaded in lanes 1 and 8, with the DNA fragment lengths 3, 2, 0.5 and 0.4 kb being pointed out. The AA281296 DNA in pT7T3D was digested with a restriction enzyme Eco RI/Not I (lane 3), Pst I (lane 6) and Xho 1 (lane 7).
30 Undigested AA281296 DNA in pT7T3D was loaded onto lane 2. 1/10 of a PCR

mixture (1 minute 94°C, 2 minutes at 60°C, 3 minutes at 72°C) with the hTC cDNA in pT7T3D and primers 1 (5' GAGTGTGTACGTC-GTCGAGCTGCTCAGGTC 3') and 4 (5' CACCCTCGAGGTGAGACGCTCGGCC 3') [lane 4] and, especially, with primers 6 (5' GCTCGTAGTTGAGCACGCTGAACAGTG 3') and 7 (5' GCCAAGTTCCTGCACTGGCTGATGAG 3') [lane 5] was applied to lanes 4 and 5.

Fig. 4: Detail from a comparison of the protein sequences of the *Euplotes* p123 (p123) and human (phTC) catalytic telomerase subunits.

The conditions (ktuple, gap penalty and gap length penalty) are listed for the Lipman-Pearson protein comparison, using the Lasergene program software (Dnastar, Inc.), which is depicted in this figure. The amino acid residues are depicted in accordance with their single-letter code. The amino acids which are identical between *Euplotes aediculatus* p123 and the identified EST$_{+1}$ are also highlighted using the corresponding letter from the single-letter code. Amino acids which are not identical but whose function is similar or comparable are marked by a :.

Fig. 5: Part of a comparison of the protein sequences of the catalytic telomerase subunits of *Euplotes* p123 (p123), and yeast (est2p).

The condition (Ktuple, gap penalty and gap length penalty) are listed for the Lipman-Pearson protein comparison using Lasergene program software (Dnastar, Inc.) which is dipicted in this figure. The amino acid residues are shown in accordance with their single letter code. The amino acids which are identical between *Euplotes aediculatus* p123 and yeast est2p are likewise given prominence by the corresponding letter from the single-letter code. Amino acids which are not identical, but which are similar or comparable in function, are marked with a :.

Fig. 6: Detail from a comparison of the protein sequences of the yeast (est2p) and human (phTC) catalytic telomerase subunits.

The conditions (ktuple, gap penalty and gap length penalty) are listed for the Lipman-Pearson protein comparison, using the Lasergene program software (Dnastar, Inc.), which is depicted in this figure. The amino acid residues are depicted in accordance with their single-letter code. The amino acids which are identical between yeast est2p and the identified EST_{+1} are also highlighted using the corresponding letter from the single-letter code. Amino acids which are not identical but whose function is similar or comparable are marked by a :.

Fig. 7:   Detail from a comparison of the protein sequences of the *Euplotes* p123 (p123), yeast (est2p) and human (phTC) catalytic telomerase subunits. The comparison, depicted in Fig. 5, between *Euplotes* p123 (p123), yeast (est2p) and humans (phTC) was carried out using the Clustal Method subprogram of the Lasergene program software (Dnastar, Inc.) under standard conditions. The amino acid residues are depicted in accordance with their single-letter code. The amino acids which are identical between yeast est2p, *Euplotes aediculatus* p123 and the identified EST_{+1} are also highlighted using the corresponding letter from the single-letter code. In addition, the regions which are identical between all three proteins are marked by a light grey bar above the protein sequence.

Fig. 8:   Generated DNA sequence from Example 6 (RACE round 1) (SEQ ID No. 3).

Fig. 9:   Generated DNA sequence from Example 6 (RACE round 2) (SEQ ID No. 4).

Fig. 10:   Generated DNA sequence from Example 6 (RACE round 3) (SEQ ID No. 5).

Fig. 11:   Generated DNA sequence from Example 8 (RACE round 3) (SEQ ID No. 6).

Fig. 12:   Outline of the cloning of the complete hTC cDNA sequence. The positions of the start and stop codons are marked by arrows. The black regions of the rectangles symbolize protein-encoding sequence sections, whereas the pale grey regions symbolize 5'- and 3'- untranslated cDNA regions and/or denote intronsequences.

The dark grey blocks in the rectangle for the full-length cDNA either denote the telomerase-specific motif (T) or the seven reverse transcriptase motifs (numbers 1-7).

The DNA fragments which are required for preparing the complete hTC cDNA are likewise depicted as rectangles and are marked in accordance with their origin. All the rectangles are arranged in their positions relative to each other. The origin of the DNA fragment which is denoted by rectangle AA261296 is described in Example 2. The relative position of the 182 bp deletion in this fragment (compare Example 2) is shown by a gap in the rectangle. The origin of the DNA fragments which are denoted by the rectangles RACE 1, RACE 2 and RACE 3 is described in Example 6. The origin of the DNA fragment which is denoted by the C5F fragment rectangle is described in Example 7. The origin of the DNA fragment which is denoted by the lambda 12 rectangle is described in Example 9. The 3' part in the lambda 12 DNA fragment which encodes a cDNA which is not connected to hTC (compare Example 9) is not depicted in this figure. The complete hTC-cDNA sequence was joined together at the 5' and 3' splice sites using the lambda 12 and C5F DNA fragments shown in this figure (compare Example 7). These splice sites were identified in a variety of fragments (RACE 1, RACE 3, lambda 12 and C5F).

Fig. 13: Detailed sections from a comparison of the protein sequences of the catalytic telomerase subunits of *Euplotes* and man (hTC).

The figure shows sections from a comparison of the protein sequences of the catalytic telomerase subunits of *Euplotes* and man (hTC). Attention is drawn to the reverse transcriptase motifs in the boxed-in areas. The figures under the boxes refer to the respective amino acid positions in Fig. 2. The amino acid residues are shown in accordance with their single-letter code. Identical amino acids are printed in bold. In the consensus sequence for the reverse transcriptase (RT consensus) motif, h denotes a hydrophobic amino acid and p denotes a polar amino acid. If these groups of amino acids are retained in the *Euplotes* and hTC amino acid sequences, p and/or h is/are then printed in bold. Very highly conserved amino acids are underlaid in grey. In RT3, the boxed-in area is extended in order to cover

additional homologous amino acids. The telomerase-specific motif is described in Example 9.

Fig. 14: Generated DNA sequence from Example 11 (3' version) (SEQ ID No. 7). The
5   region which is not homologous with the DNA sequence depicted in Fig. 1 is made
          to stand out in bold.

Fig. 15: hTC expression in cancer cell lines and normal human tissue. Fig. A:
          Approximately 2 µg of poly-A⁺ RNA from different human cell lines were
10   immobilized on the Northern blot in accordance with the manufacturer's
          (Clontech) instruction. Specifically, the RNA originated from a melanoma (G361),
          a lung carcinoma (A549), an adenocarcinoma of the colon (SW480), from a Raji
          Burkitt's lymphoma, from a leukaemia cell line (MOLT-4), from a chronic
          leukaemia cell line (K-562), from a cervical tumour (HeLa) and from the
15   leukaemia cell line HL60. The transcripts marked 4.4 kb, 6 kb and 9.5 kb are
          specific for hTC (compare Example 10). Fig. B: About 2 µg of poly-A⁺ RNA from
          different human tissues were immobilized on the Northern blot in accordance with
          the manufacturer's (Clontech) instructions. Specifically, the RNA was isolated
          from heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas. An
20   RNA size standard is shown.

Fig. 16: Western blot analysis of the rabbit sera against peptides from the human
          telomerase amino acid sequence (Example 12). In each case, 20 µl of the bacterial
          lysates from Example 13 were analysed in a western blot (Ausubel *et al.*, 1987)
25   using the antisera from Example 12. Lysates from bacteria which harbour the
          pMALEST construct were loaded in lanes 1, 2, 6 and 7. Lysates from bacteria
          which harbour the pMALA1 construct were loaded in lanes 3, 4, 8 and 9. Lysates
          from bacteria which were not induced with IPTG (isopropyl-beta-
          thiogalactopyranoside) were loaded in lanes 1, 3, 6 and 8. Lysates from IPTG-
30   induced bacteria were loaded in lanes 2, 4, 7 and 9. A standard size marker
          (10 kDa protein ladder from Life Technologies, Cat. No. 10064-012) was loaded

in lane 5. The 50 kDa and 120 kDa bands are marked at the edges of the membranes. The PVDF membrane in Fig. A containing lanes 1 to 4 was incubated with preimmune sera against peptide B (compare Example 12). The PVDF membrane in Fig. B containing lanes 6 to 9 was incubated with preimmune sera against peptide C (compare Example 12). The PVDF membrane in Fig. B containing lanes 1 to 4 was incubated with immune sera against peptide B (compare Example 12). The PVDF membrane in Fig. B containing lanes 6 to 9 was incubated with immune sera against peptide C (compare Example 12).

Fig. 17:   Autoradiogram of ^{35}S-labelled, *in vitro*-translated protein. The complete *in vitro*-translated hTC (compare Example 15) was loaded in lane 1. A C-terminally truncated version of phTC was loaded in lane 2. Lane 3 shows a positive control for the *in vitro* translation which was supplied by the manufacturer (compare Example 15). A protein size standard for estimating protein sizes is marked on the right-hand side.

Fig. 18:   Autoradiogram of ^{32}P-labelled products from the TRAP assay (compare Example 15). A TRAP assay mixture without any added enzyme or protein was loaded, as a negative control, in lanes 1 and 2. A TRAP assay mixture containing partially purified human telomerase from HeLa cells was loaded, as a positive control, in lanes 3 and 4. A TRAP assay mixture containing *in vitro*-translated phTC was loaded, undiluted, in lanes 5 and 6. A TRAP assay mixture containing *in vitro*-translated phTC, at a 1:4 dilution, was loaded in lanes 7 and 8. A TRAP assay mixture containing *in vitro*-translated phTC, at a dilution of 1:16, was loaded in lanes 9 and 10. A TRAP assay mixture containing *in vitro*-translated luciferase was loaded, as a negative control, in lanes 11 and 12.

Fig. 19:   Autoradiogram of ^{32}P-labelled products from the direct telomerase assay (compare Example 15). A radioactively labelled 10 bp marker was loaded in lane 1. A telomer oligonucleotide ([TTAGGG]₃) which was radioactively labelled 5' was loaded in lane 2. Lane 3 is an empty lane. Partially purified human telomerase

from HeLa cells was used in a direct assay and the synthesis product was loaded, as a positive control, in lane 4. The *in vitro*-translated phTC from Example 15 was used in a direct assay and the synthesis product was loaded in lane 5.

Examples

Example 1

It is nowadays accepted that less than 5% of the human genome is in fact transcribed and translated into protein. Even before the genome has been completely sequenced, it is possible to obtain important information about the 60,000-70,000 genes in a human cell by investigating these coding moieties of the genome in a specific manner. The automation of high-throughput DNA sequencing technology in the last 10 to 15 years has made it possible to collect many cDNAs from plasmid cDNA libraries of widely differing origin and sequence the 5' or 3' end in each case. These short DNA sequences, which are typically of from 300 to 400 bp in length, are termed expressed sequence tags or ESTs for short and are compiled in various specialized data bases. The EST approach was initially described by Okubo et al. (1992) and transferred to a larger scale by Adams et al. (1992). At present, approximately 50,000 human cell genes are partially sequenced and documented as EST entries.

By comparing with the DNA and amino acid sequences of known genes, it is possible to identify related, but hitherto unkown, genes in these EST databases (Gerhold and Caskey, 1996). tBLASTn (Altschul et al., 1990) is a search algorithm which has proved particularly useful for this purpose. This algorithm translates every DNA clone in the EST data base in all six possible reading frames and compares these amino acid sequences with the known protein sequence.

The EST data base at the National Center for Biotechnology Information (NCBI) was searched with the recently published protein sequence for the *Euplotes aediculatus* catalytic telomerase subunit *p123* (Lingner et al., 1997). This resulted in a human EST with the accession number AA281296 being identified which exhibits significant homology with p123 in reading frame +1. This amino acid sequence in reading frame +1 is termed Est_{+1} in that which follows.

The homology between p123 and the Est$_{+1}$ is most conspicuous in two sequence regions which are separated by 30 amino acids. The longer sequence region, which in p123 extends from amino acid 438 to amino acid 484, is 38% identical to the corresponding region Est$_{+1}$. If similar amino acids are also taken into consideration, the congruence is even 59%. The second block of homology extends, in the p123 protein, from amino acid 513 to amino acid 530 and exhibits 44% identity with the corresponding sequence segment in the identified Est$_{+1}$. A congruence of 61% is obtained when amino acid residues having similar properties are taken into account.

The P (probability) value is an important parameter for assessing a BLAST search. P indicates the probability of also finding a specific segment pair in a BLAST search using a random sequence and varies numerically between 0 (highly significant result) and 1 (insignificant result). Thus, comparison of the p123 equivalent from yeast (est2p) with the NCBI EST data base, for example, gave a negative result: The EST which was found had a probability of P=1 (Tab. 1). On the other hand, human telomerase-associated protein 1 (hTP1), which was found in an EST data base which is not available to the general public (Harrington et al., 1997), gives a probability of P=0.004.

known gene (species)	P	identified gene	origin of the cDNA library
est2p (*Saccharomyces cerevisiae*	0.999	Rat EST	Kidney
p80 (*Tetrahymena termophilia*)	0.004	hTP1 (Harrington *et al.*, 1997)	Crypts of the intestinal epithelium
p123 (*Euplotes aediculatus*)	$3.5^x 10^{-06}$	AA281296	Germinal centres of the tonsils

Tab. 1:   Comparison of three tBlastn search runs using different known genes.

- 23 -

The human EST AA281296 which was identified by the comparison with p123 has a probability of $P=3.5 \times 10^{-6}$.

These data suggest that the identified EST in all probability encodes a fragment of the

5   catalytic subunit of human telomerase. For this reason, the corresponding gene is abbreviated below to hTC (human Telomerase, catalytic) and the deduced protein is abbreviated to phTC.

Example 2

The EST which was identified by the comparison with p123 was fed into the EST data base

10   on 2 April 1997 and has not been published in any journal. According to information obtained from the National Center for Biotechnology Information, the cDNA library which contains this EST clone was prepared as follows:

After the mRNA had been prepared from the germinal centres of the tonsils, a cDNA

15   synthesis was carried out and the double-stranded cDNA fragments were cloned in an orientated manner, using the Not I and Eco RI restriction enzyme cleavage sites, into the vector pT7T3D-Pac.

The 389 bp which had been fed into the EST database were sequenced using the -28m13 rev2

20   primer supplied by Amersham (DNA sequence, see Fig. 1 Position 1685 to 2073).

Lasergene program software (Dnastar Inc.) was used to translate the DNA sequence of EST AA281296 in accordance with the human genetic code. The resulting amino acid

25   sequence (Est$_{+1}$) corresponds to Position 542 to 670 in Fig. 2.

The deduced protein sequence of Est$_{+1}$ is composed of 129 amino acids, including 27 basic, 11 acidic, 51 hydrophobic and 28 polar amino acid residues.

The EST (AA281296) which was identified in Example 1 was obtained commercially from Research Genetics, Inc. (Huntsville) in the form of a plasmid transformed into *E. coli* and analyse experimentally:

5   As shown in the ethidium bromide-stained agarose gel depicted in Fig. 3, a fragment from EST AA281296 of approximately 2.2 kb in size is liberated from the vector pT7T3D after subjecting the prepared plasmid DNA to restriction digestion. With the aid of a polymerase chain reaction (PCR), which was carried out in parallel and which made use of specific internal primers, EST AA281296 was inspected: the lengths of the expected PCR products

10   are 325 and 380 bp and are in agreement with the lengths of the fragments which were found experimentally (cf. tracks 4 and 5 in Fig. 3). This therefore demonstrated that the E.coli clone supplied by Research Genetics, Int. (Huntsville) therefore harbours the identified EST as a plasmid.

15   After the DNA had been prepared, the 2176 bp of the insert in total were identified by means of double-strand sequencing. A comparison of the DNA sequences of clone AA281296 and of the C5F fragment (compare Example 7) showed that there was a 182 bp deletion (Positions 2352 to 2533, Fig. 1) and that the open reading frame is consequently displaced in this region. In summary, the DNA sequence of clone AA281296 is composed of the sequence

20   information shown in Fig. 1 (Positions 1685 to 2351 and Positions 2534 to 4042).

Example 3

The tBLASTn comparison only identifies the regions in which there is the greatest agreement

25   between p123 and Est_{+1} (amino acids 438-530, in p123), whereas the intervening amino acids are not taken into account. A Lipman-Pearson protein comparison was carried out in order to be able to draw conclusions about the relatedness of the protein sequences over a larger region (amino acids 437-554, in p123) (see Fig. 4). When this was done, 34% of the amino acids were found to be identical while 59% of the amino acids were found to be either

30   identical or biochemically similar. This result demonstrates that the relatedness of these

proteins also continues outside the regions of homology which were found using the tBLASTn program.

As has recently been reported (Lingner *et al.*, 1997), *Euplotes aediculatus* p123 and *Saccharomyces cerevisiae* est2p are homologous to each other. In order to relate the degree of affinity between p123 and est2p to the homology between p123 and Est_{+1} which is described here, the Lipman-Pearson protein comparison was employed to compare the above-described region of p123 (amino acids 437-554) with est2p, too, using identical parameters. This showed that, in this chosen region, p123 and est2p are 21% identical and that 22% of their amino acid residues are either identical or biochemically similar (see Fig. 5). Accordingly, the homology between Est_{+1} and *Euplotes* p123 is significantly higher than between p123 and est2p.

Example 4

The homology of p123 with Est_{+1} and est2p suggests that all 3 proteins belong to the same protein family. In order to confirm this assumption, est2p was compared with Est_{+1} under the conditions described in Example 3 (see Fig. 6). This showed that Est_{+1} is 20% identical to est2p, that is exhibits a degree of homology which is comparable to that of p123 to est2p. This comparatively low level of congruence also confirms the finding that no significant EST was identified in the tBLASTn search using est2p (see Example 1).

Example 5

A computer comparison using p123, est2p and phTC was carried out in order to identify possibly functional domains which are important for the protein family consisting of catalytic telomerase subunits derived from different species (see Fig. 7). In this analysis, two regions which are present in all three proteins are particularly conspicuous (see Fig. 7). At present, no unambiguous function can be assigned to the region which, in p123, corresponds to amino acids 447 to 460 (Fig. 13, telomerase motif). A motif search using the Genetics Computer

Group (GCG) Wisconsin Sequence Analysis Package and a search in a protein data base (Swissprot, version of 8.6.1997) did not provide any significant insights.

On the other hand, a second region which is homologous between p123, est2p and phTC, corresponding in p123 to amino acids 512-526, exhibits a consensus motif for a reverse transcriptase (RT) (Figs. 7 and 13). Lingner *et al.*, (1997) showed that p123/est2p contain a total of 6 such RT motifs, which are essential for the catalytic function of p123/est2p. As depicted in Figs. 7 and 13, two such RT motifs are also conserved in the sequence of phTC which has been investigated. These motifs are the RT motifs which are located to the furthest extent N-terminally in p123/est2p (Lingner *et al.*, 1997).

The primary sequences of reverse transcriptases are strongly divergent; only a few amino acids are fully conserved within a separate motif (Poch *et al.*, 1989 and Xiong and Eickbush, 1990). In addition, due to having different distances between the conserved RT motifs, reverse transcriptases which are encoded by retroviruses or long terminal repeat (LTR) retroposons differ from those reverse transcriptases which are encoded by non-LTR retroposons or group II introns (Xiong and Eickbush, 1990). Based on the structure of their RT motifs, p123, est2p and phTC are to be assigned to the latter RT group. Interestingly, in this context, the consensus sequences of the RT motifs in phTC correspond most closely to the postulated RT consensus motif: of eight amino acid residues within the two RT motifs, 6 are present in the case of phTC while only 5 are present in the case of p123 and esp2p (Figs. 7 and 13). It is striking in this context how the hydrophobic amino acids, such as leucine and isoleucine, and the amino acids lysine and arginine, in particular, are in specific positions (Figs. 7 and 13).

In summary, it was hereby possible to demonstrate, at the descriptive level, that the AA281296 clone, identified due to its homology with p123, is a fragment of the catalytic subunit of human telomerase.

Example 6

For cloning the 5' end of the hTC-cDNA, three consecutive RACE (rapid amplification of cDNA ends) reactions were carried out in addition to the homology screening described in Example 8. Marathon-Ready cDNA (Clontech) form the human leukaemia cell line K562 or from human testis tissue was employed as the cDNA source. The implementation of the individual RACE rounds, as well as the results obtained, are described below.

In addition to this, the sequence information obtained in the RACE rounds was used in order to amplify the individual fragments from a contiguous cDNA clone by means of PCR.

RACE round 1:

In a final volume of 50 µl, 10 pmol of dNTP-mix were added to 5 µl of K562 Marathon-Ready cDNA (from Clontech, Catalogue Number 7441-1), and a PCR reaction was carried out in 1 × Klen Taq PCR reaction buffer and 1 × advantage Klen Taq polymerase mix (from Clontec). 10 pmol of the internal gene-specific primer GSP2 (5'-GCAACTTGCTCCAGACACTTCTTCCGG-3') from the 5' region of the hTC-EST clone and 10 pmol of the Marathon Adaptor primer AP1 (5'-CCATCCTAATACGACTCACTATAGGGC-3'; from Clontech) were added as primers. The PCR was carried out in 4 steps. After a one-minute denaturation at 94°C, denaturation was then carried out for 5 cycles of 30 sec at 94°C and the primers were then subsequently annealed for 4 min at 72°C and the DNA chain was extended. There then followed 5 cycles in which the DNA was denatured for 30 sec at 94°C but the subsequent primer extension took place for 4 min at 70°C. Finally, 22 cycles were then carried out in which, after the 30 sec DNA denaturation, the primer annealing and chain extension took place for 4 min at 68°C.

Following this PCR, the PCR product was diluted 1:50. 5 µl of this dilution were used in a second "nested" PCR together with 10 pmol of dNTP-mix in 1 × 10 Klen Taq PCR reaction buffer and 1 × Advantage Klen Taq polymerase mix and also 10 pmol of primer GSP2 and

10 pmol of the "nested" Marathon Adaptor primer AP2 (5'-ACTCACTATAGGGCTCGAGCGGC-3'; from Clontech). The PCR conditions corresponded to the parameters selected in the first PCR. As the only exception, only 16 cycles were chosen, instead of 22 cycles, in the last PCR step.

5

A DNA fragment of 1153 bp in length was obtained as the product of this nested RACE PCR. This fragment was cloned into the TA cloning vector pCR2.1 from Invitrogen and subjected to complete double-strand sequencing (Fig. 8 and SEQ ID No. 3).

10 Nucleotides 974 to 1153 represent the nucleotide region 1629 to 1808 of the hTC-cDNA which is depicted in Fig. 1. The nucleotide region extending from bp 1 to bp 973, which does not exhibit any homology with the hTC-cDNA sequence shown in Fig. 1, represents intron sequences of the hTC gene (data not shown). A 3' splice consensus sequence is located at the exon-intron transition. The presence of intron sequences could be due to using incompletely

15 spliced mRNA as the starting substance for the cDNA synthesis. Genomic DNA contamination in the cDNA could also be an explanation for intron sequences being found.

RACE round 2:

20 Based on the sequence data obtained in the first RACE round, a second RACE was carried out using the gene-specific primer GSP5 from the 5' region of RACE product 1 (5'-GGCAGTGACCAGGAGGCAACGAGAGG-3') and the AP1 primer. Marathon-Ready cDNA from human testis (from Clontech; Catalogue Number 7414-1) was used as the cDNA source. The same PCR conditions were selected as in the 1st PCT in RACE round 1. The 1st

25 PCR was also followed, in RACE round 2, by a 2nd "nested" PCR using diluted PCR product as the cDNA source. The gene-specific primer GSP6 from the 5' region of RACE product 1 (5'-GGCACACTCGGCAGGAAACGCACATGG-3') and the AP2 primer were used as the "nested" PCR primers. The conditions corresponded to parameters for the nested PCR from RACE round 1.

30

The PCR product of 412 bp in length from the nested PCR of RACE round 2 was cloned into the TA cloning vector pCRII-Topo from Invitrogen and sequence completely (Fig. 9 and SEQ ID No. 4). The sequence segment from bp 267 to bp 412 is completely homologous with the 5' region of the product from RACE 1. The region from bp 1 to bp 266 extends RACE product 1 at the 5' end. This RACE product 2 is probably, in its entirety, an intron region of the hTC gene (data not shown).

RACE round 3:

A third RACE round led to the identification of hTC-cDNA regions which were located further on in the 5' direction. Using the sequence results from RACE round 2 as a base, a gene-specific primer GSP9 (5'-CCTCCTCTGTTCACTGCTCTGGCC-3') was selected from the 5' region of RACE product 2 and used in a new RACE together with the AP1 primer and Marathon-Ready cDNA from human testis (from Clontech). The RACE conditions were the same as those used in the 1st PCR in RACEs 1 and 2. In the "nested" RACE which followed, and which took place, in accordance with the "nested" RACEs in rounds 1 and 2, using the gene-specific primer GSP10 from the 5' region of RACE product 2 (5'-CGTAAGTTTATGCAAACTGGACAGG-3') and AP2, a fragment of 1012 bp in length (Fig. 10 and SEQ ID No. 5) was amplified and cloned into the TA cloning vector pCRII--TOPO. Subsequent sequencing showed that the 3' region of this RACE fragment (bp 817 - bp 1012) evidently still constitutes an intron sequence of the hTC gene. The region from bp 889 to bp 1012 is completely homologous with the 5' region of RACE product 2. On the other hand, the 5' region of this fragment, from bp 1 to bp 816, is identical to the bp 814 - bp 1629 region of the hTC-cDNA which is shown in Fig. 1. A potential 5' splice consensus sequence is located at the exon-intron transition.

Example 7

A PCR was carried out in order to clone a contiguous fragment from the sequence information obtained from RACE 2 and clone AA281296. Marathon-Ready cDNA from human testis (from Clontech; Catalogue Number 7414-1) was used as the cDNA source. The

PCR mixture was as described under RACE 1 (compare Example 6) but using the primers C5F (5'-CGAGTGGACACGGTGATCTCTGCC-3') from the 5' region of RACE 2 and primer C3B (5'-GCACACCTTTGGTCACTCCAAATTCC-3') from a 3' region of clone AA281296. The PCR was carried out in 2 steps. After a one-minute denaturation at 94°C, denaturation was then carried out for 36 cycles of 30 sec at 94°C and, after that, the primers were annealed, and the DNA chain was extended, for 4 min at 68°C.

A DNA fragment of 2486 bp in length, which is designated the C5F fragment below, was obtained as the product of this PCR. This fragment was cloned into the TA cloning vector pCRII-TOPO from Invitrogen and subjected to complete double-strand sequencing. A comparison of the DNA sequences of the C5F fragment and the AA281296 clone showed that there was an in-frame insertion of 182 bp between RT motif 3 and RT motif 4 (Positions 2352 to 2533, Fig. 1). A further comparison of DNA of the C5F fragment with the sequences from the three RACE rounds made it clear that an intron which was already identified in RACE 2 was present at the 3' end of C5F. A 3' splice consensus sequence is located at the exon-intron transition. In summary, the DNA sequence of the C5F fragment is consequently composed of the sequence information shown in Fig. 9 (Position 64 to 278) and the sequence data shown in Fig. 1 (Positions 1636 to 3908).

Example 8

For cloning the 5' end of the hTC-cDNA, a homology screening (Ausubel *et al.*, 1987) was carried out in addition to the RACE protocol described in Example 6. A human erythroleukaemia 5'-stretch plus cDNA library (from Clontech, cat. No. HL5016b) from the human leukaemia cell line K562 was used as the cDNA source. Approximately 3×10^6 Pfu of this random and oligo-dT-primed library were plated out and used for screening as described in Ausubel *et al.* (1987). A radioactively labelled hTC-DNA fragment of 719 bp in length (Positions 1685 to 2404, corresponding to Fig. 1) was used as the probe.

Following a rescreening with the same hTC probe, the λ clone 12 was verified as being positive out of 20 putatively positive λ clones. Following plaque purification and λ DNA

preparation (Ausubel *et al.*, 1987), the 4 kb insert was recloned into the pBluescript vector and sequenced (Fig. 11 and SEQ ID No. 6).

A comparison of the λ clone 12 sequence with the sequences of the RACE clones and the

5   DNA sequence of clone AA281296 showed that this clone, which was identified in the homology screening, encodes a 5' part of the hTC-cDNA and possesses a putative ATG start codon in Position 63 in accordance with Fig. 1. There is no stop codon in the same reading frame 5' of this ATG. Subsequent sequence analyses make it clear that λ clone 12 probably contains an intron from Positions 1656 to 2004. Very well conserved 5' and 3' splice sites

10   provide support for this hypothesis. The hTC-cDNA-encoding sequence then continues from Position 2005 to Position 2382. The sequence from 2383 to the 3' end of λ clone 12 exhibits a conspicuous open reading frame in reading frame -4. A bioinformatic analysis of the corresponding DNA sequence showed that, over about 400 bp, this reading frame is identical to a variety of ESTs which have no connection with the hTC cDNA. Consequently,

15   λ clone 12 is a chimeric clone which essentially consists of the 5' end of the hTC cDNA and another cDNA clone of unknown function.

A diagrammatic summary showing the relative orientations of the RACE products, and the homology screening, is depicted in Fig.12. The complete sequence of the hTC cDNA (Fig. 1)

20   was assembled from λ clone 12 (Positions 21 to 1655 in accordance with Fig. 11), the C5F PCR product (Positions 1636 to 3908 in accordance with Fig. 1) and EST AA281296 (Positions 3909 to 4042, in accordance with Fig. 1).

Example 9

25

A total of seven motifs for reverse transcriptases (RT motifs) was identified by comparing the phTG protein sequence (Fig. 2 and SEQ ID NO. 2) with a reverse transcriptase consensus sequence (Poch *et al.*, 1989, Xiong and Eickbush, 1990) (Fig. 13). Within these motifs, some amino acids are highly conserved not only between the RT consensus sequence and phTC but

30   also in comparison with the *Euplotes* telomerase protein. Thus, two aspartic acids (Positions 868 and 869 in Fig. 2) are, for example, completely conserved in RT motif 5 (Fig. 13). RT

motif 7, which was deduced from other reverse transcriptases (Poch *et al.*, 1989, Xiong and Eickbush, 1990), was only demonstrated in the human catalytic telomerase subunit and not in the *Euplotes* protein (Fig. 13).

5   Structural features which can only be found in the telomerase proteins and not in other reverse transcriptases are also conspicuous. The telomerase motif (Positions 553 and 565 in Fig. 2) is a structure which is specific for this protein family since it does not occur in any previously known protein. A further feature which has only been identified in the catalytic telomerase proteins is the difference between RT motifs 3 and 4, which distance, at 10   107 amino acids, is markedly greater than in other RTs. These special features indicate that the catalytic subunits of the telomerases from different species probably constitute a separate subgroup of RNA-dependent DNA polymerases.

Example 10

15   Expression of the telomerase RNA subunit (hTR) does not correlate with telomerase activity but, instead, is observed ubiquitously (Feng *et al.*, 1995). Consequently, the question arises as to whether expression of the catalytic telomerase subunit is associated with telomerase activity.

20   Northern blot experiments (Ausubel *et al.*, 1987) were carried out in order to analyze the level of hTC expression. The commercially available Northern blots were supplied with a number of RNA preparations from normal human tissue (from Clontech; catalogue No. 7760-1) or with RNA samples from human cancer cell lines (from Clontech; Catalogue 25   Number 7757-1). A radioactively labelled hTC DNA fragment of 719 bp in length (Positions 1685 to 2404, in accordance with Fig. 1) was used as the probe. The membranes were incubated with the probe in accordance with the manufacturer's (Clontech) instructions.

Two main RNA transcripts, of about 9.5 kb and 4.4 kb in size, and an additional RNA 30   transcript of about 6 kb, which transcripts cross-hybridize with the probe, were detected in the eight human cell lines (3 leukaemia cell lines, 3 carcinoma cell lines, one melanoma and one

lymphoma) tested (Fig.15, Fig. A). In the comparison, the hTC mRNA was expressed most strongly in the leukaemia cell lines K-562 and HL-60 (Fig. 15, Fig. A). By contrast, it was not possible to detect the hTC transcript in the normal tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney and pancreas) which were tested (Fig. 15, Fig. B). This observation is not surprising since it was not possible to detect any telomerase activity, either, in these tissues (Kim *et al.*, 1994).

These data indicate that the induction of hTC expression plays an important role in activating the telomerase during tumour development.

Example 11

Several PCR products, whose sizes only differed from each other to a minimal extent, were always obtained when the hTC cDNA fragments from various cDNA libraries (Clontech Marathon Ready cDNA from the human leukaemia cell line K562 and from human testis and also cDNA from the human premyeloid leukaemia cell line HL60) were subjected to PCR amplification. In order to elucidate the differences between the different hTC-PCR products, a fragment of the hTC cDNA depicted in Fig. 1 extending from bp 1783 to bp 3901 was amplified using the primers C5A (5'-CCGGAAGAGTGTCTGGAGCAAGTTGC-3') and C3B (5'-GCACACCTTTGGTCACTCCAAATTCC-3'). Marathon-Ready cDNA from K562 leukaemia cells (from Clontech; Catalogue Number 7441-1) was used as the cDNA source (PCR1 and 2). In a third PCR, a hTC fragment, from bp 1695 to bp 3463, of the hTC cDNA in Fig. 1 was amplified from HL60 cDNA using the primers GSP1 front (5'-GGCTGATGAGTGTGTACGTCGTCGAG-3') and HTRT3A (5'-GGGTGGCCATCAGTCCAGGATGG-3').

The conditions of the 3 PCR reactions are described below:

In the first PCR, and in a final volume of 50 μl, 10 pmol of dNTP mix were added to 5 μl of K562 Marathon-Ready cDNA, and a PCR reaction was carried out in 1 × Klen Taq PCR reaction buffer and 1 × Advantage Klen Taq polymerase mix (from Clontech). 10 pmol of

each of the primers C5A and C5B were added. The PCR was carried out in 3 steps. A one-minute denaturation at 94°C was followed by 35 PCR cycles in which the DNA was firstly denatured for 30 sec at 94°C and the primers were then annealed, and the DNA chain was extended, for 4 min at 68°C. In conclusion, there followed a chain extension for 10 min

5   at 68°C. The resulting PCR products were cloned into the TA cloning vector pCRII-TOPO from Invitrogen.

In a second PCR, 10 pmol of each of the primers C5A and C3B, 10 pmol of dNTP mix and 2 U of Taq DNA polymerase (from Gibco-BRL) were added to 5 µl of K562 Marathon-

10   Ready cDNA, and a PCR reaction was carried out in 1 × PCR buffer (from Perkin Elmer) in a final volume of 50 µl. The PCR reaction was carried out in 3 steps. The DNA was firstly denatured for 3 min at 94°C. There then followed 34 cycles in which, consecutively, the DNA was denatured for 45 sec at 94°C, primer annealing then took place for 1 min at 68°C and, after that, the DNA chain was extended for 3 min at 72°C. In the last PCR step, a

15   concluding chain extension was carried out for 10 min at 72°C. The resulting PCR products were cloned into the TA cloning vector pCR2.1 from Invitrogen.

For the third PCR, the cDNA synthesis kit from Boehringer Mannheim was first of all used to carry out a cDNA synthesis from 2 µg of DNaseI-treated poly-A RNA from the human premyeloid cell line HL60 in accordance with the manufacturer's instructions. 1 µl of this

20   HL60 cDNA was then mixed with 10 pmol of each of the primers GSP1 front and HTRT3A and also 10 pmol of dNTP mix, in a final volume of 50 µl, and, after 1.25 µl of DMSO in 1 × Klen Taq PCR reaction buffer and 1 × Advantage Klen Taq polymerase mix (from Clontech) had been added, a PCR reaction was carried out. The PCR reaction proceeded in

25   3 steps. After a denaturation for 3 min at 94°C, the DNA was initially denatured for 1 min at 94°C and the primers were then annealed, and the DNA chain extended, for 4 min at 68°C, over 37 cycles. The reaction was concluded by a further incubation for 10 min at 68°C. The PCR products were cloned into the TA cloning vector pCR2.1-TOPO.

Complete double-strand sequencing of the cloned hTC cDNA fragments from PCRs 1 and 2, and partial sequencing of the hTC cDNA fragments obtained from PCR 3, showed that, in addition to the hTC cDNA depicted in Fig. 1, 4 variants of this cDNA exist in human cells, i.e.:

Variant 1 of human hTC cDNA is distinguished by a deletion of 182 bp in length extending from nucleotides 2345 to 2526. This deletion results in the ORF being displaced, with a truncated hTC protein, which lacks RT motifs 4 to 7, being read off.

10   Variant 2 of human hTC cDNA exhibits a deletion of 36 bp in length extending from nucleotides 2184 to 2219. RT motif 3 is lost as a result of this deletion. However, the reading frame is retained and a protein is produced which selectively lacks RT motif 3.

Variant 3 of human hTC cDNA is a combination of variants 1 and 2. It exhibits both a

15   deletion from bp 2184 to 2219 and a deletion from bp 2345 to 2526.

Variant 4 of human hTC cDNA is distinguished by the loss of the nucleotide region from bp 3219 to bp 3842. This missing sequence is replaced by a sequence which is not homologous with hTC. From bp 3843 onwards, the sequence is once again completely

20   identical to the hTC sequence depicted in Fig. 1. The sequence of variant 4 is shown in Fig. 14. In accordance with the 5' primer chosen, it begins with bp 1783 of the hTC cDNA shown in Fig. 1. The region which is not homologous is emphasized in bold and, from Position 3219 to Position 3451 (Fig. 14 and SEQ ID No. 7) is, to the extent of 98.7%, in agreement, at the DNA level, with an EST (Accession No. AA299878) from a human uterus

25   tumour.

Exanple 12

In order to obtain antisera having specificity for the catalytic subunit of human telomerase,

30   the available nucleotide sequence (Fig. 1) was translated into an amino acid sequence (Fig. 2). Using a secondary structure prediction program (PROTEAN, from the DNAStar

software package, DNASTAR Inc., Madison, WI, USA), two peptides were chosen which, with a certain degree of probability, evoke an immune response. These are the following peptides, which are depicted in the one-letter code for amino acids:

5　　　B:　　　C-K-R-V-Q-L-R-E-L-S-E-A-E-V-R-Q - $CONH_2$/Pos. 594 - 608

　　　　　C:　　　C-Q-E-T-S-P-L-R-D-A-V-V-I-E-Q-S-S-S-L-N-E - $CONH_2$/Pos. 781-800

The cysteines which are underlined are not derived from the telomerase sequence but were additionally added on as linkers for the coupling.

10

The peptides were coupled to keyhole limpet hemocyanin (KLH) using the thiol-reactive coupling reagent m-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS). Two rabbits were in each case immunized with these coupled peptides at intervals of from 2 to 4 weeks. Prior to immunization, 5 ml of blood were withdrawn in order to obtain preimmune sera.
15　After 4 immunizations, 5 ml of blood were likewise withdrawn for obtaining immune sera. These sera were tested for reactivity with fusion proteins (Example 13) in a Western blotting experiment (Ausubel et al., 1987).

Example 13

20　Bacterial expression experiments were carried out in order to be able to analyse the protein of the catalytic telomerase subunit.

The constructs of these experiments are described below:

25

For the expression construct pMalEST, the insert in the AA281296 clone mentioned in Example 2 was excised with restriction enzymes Eco RI and Not I and the cleavage sites were filled in using the Klenow fragment (Ausubel et al., 1987); the insert was then cloned into the given reading frame of the maltose-binding protein of the bacterial expression vector
30　pMAL-C2 (from New England Biolabs). Vector pMAL-C2 was digested with restriction

enzyme Pst I and the protruding single-strand ends were removed with T4 DNA polymerase (Ausubel *et al.*, 1987).

The expression construct pMalA1 contains the nucleotide sequence of Fig. 1 from Position 1789 to Position 3908. This DNA fragment was amplified from a commercially available K562 Marathon-Ready cDNA library (from Clontech, Catalogue Number 7441-1) by means of PCR using the primers C5A (5'-ACCGGAAGAGTGTCTGGAGCAAGTTG-3') and C3B (5'-GCACACCTTTGGTCACTCCAAATTCC-3'), and cloned into the TA cloning vector pCRII-TOPO from Invitrogen. The PCR conditions were as described in Example 7. For the expression construct pMalA1, the insert was excised using the restriction enzyme Eco RI and the cleavage sites were filled in using the Klenow fragment (Ausubel *et al.*, 1987); the insert was then cloned into the bacterial expression vector pMAL-C2 (from New England Biolabs) which had been cleaved with the restriction enzyme Xmn I.

These constructs were then used for protein expression in the bacterial strain *E. coli* DH5α. The expression conditions were those as described in the instructions provided by New England Biolabs (Catalogue Number 800). The bacterial lysates which were prepared were tested in a Western blotting experiment (Ausubel *et al.*, 1987).

Example 14

The bacterial lysates from Example 13 were analysed in a Western blot (Ausubel *et al.*, 1987) using the antisera from Example 12.

Since the proportion of the fusion represented by the maltose-binding protein is about 43 kDa in size, fusion proteins of about 74 kDa and 106 kDa are expected for the pMalEST and pMalA1 constructs, respectively.

When comparing the preimmune sera with the sera following the first immunization, it becomes evident that specific antibodies were formed against the B and C epitopes (Fig. 16). Furthermore, in addition to the expected 74 kDa and 106 kDa proteins, respectively, smaller

protein fragments were also observed which react with the antisera. These smaller products probably originate from premature products.

Only the epitope for serum B is present on the fusion protein from the expression using pMalEST. By contrast, the epitopes for sera B and C are present on the fusion protein from pMalA1. For this reason, antiserum C does not recognize the pMalEST expression product and only recognizes the larger protein fragments from the expression experiments using pMalA1. This observation underlines the high degree of specificity of the antisera which were generated.

Example 15

In order to be able to analyse the protein of the catalytic telomerase subunit, the protein component should be reconstituted *in vitro* together with the RNA component.

The constructs for these experiments are described below:

The RNA component of 504 nt in length (Feng *et al.*, 1995) was amplified from a 293 cell cDNA library using the primers HTR9BAM (5'-CGCGG-ATCCTAATACGACTCACTATAGGGTTGCGGAGGGTGGGCCTG-3') and HTR2BAM (5'-CGCGGATCCCGGCGAGGGGTGACGGATGC-3). Primer HTR9BAM contains a T7 promoter from nucleotide 10 to 29. In the PCR, 10 pmol of dNTP mix were added, in a final volume of 100 µl, to 3 µl of cDNA from 293 cells, and a PCR reaction was carried out in 1 × PCR reaction buffer containing 0.5 µl of Taq polymerase (from Gibco). 10 pmol of each of the primers HTR9BAM and HTR2BAM were added. The PCR was carried out in 3 steps. A ten-minute denaturation at 94°C was followed by 35 PCR cycles in which the DNA was first of all denatured for one minute at 94°C and, after that, the primers were annealed, and the DNA chain was extended, for 2 min at 62°C. In conclusion, there followed a chain extension for 4 min at 72°C. The resulting PCR products were cloned, after a restriction digestion with Bam HI, into the Bam HI cleavage site of vector pUC19 in such a way that the

RNA component is under the control of the T7 promoter. This construct is designated HTR504 in that which follows.

5   The cDNA fragment of 3411 bp in length (Position 60 to Position 3470, Fig. 1) was cloned into the vector PCRII TOPO (from Invitrogen). Detailed information on the cloning is given in Examples 8 and 7, and also in Fig. 12. In this construct, which is designated HTC FL, the T7 promoter is located 5' before the hTC cDNA.

10 The catalytic telomerase protein component was synthesized in a commercially available transcription/translation system, after adding the hTC FL construct, in accordance with the manufacturer's (Promega; Catalogue Number L4610) instructions. Whether the *in vitro* translation of the expected 127 kDa product had been successful was checked in an SDS-PAGE (Ausubel *et al.*, 1987) using ^{35}S-labelled cysteine (Fig. 17).

15 The telomerase RNA component was synthesized using a transcription system in accordance with the manufacturer's (Ambion; Catalogue Number 1344) instructions or using the method described by Pokrovskaya and Gurevich (1994).

20 For the *in vitro* re-constitution, 0.5 μg of hTRNA was added to 50 μl of the above-described translation mixture containing the hTC FL construct and the whole was incubated at 37°C for 10 min. The enzymatic activity of 2 μl of this mixture was investigated using the TRAP assay (N.W. Kim *et al.*, 1994). The measurement of the activity, by the same method, of telomerase which was purified from HeLa cells (Shay *et al.*, 1994) was used as the positive control. As can be seen in Fig. 18, both the reconstituted enzyme and the native enzyme produce the

25 same product pattern, i.e. the nucleotide ladder which is characteristic for telomerase. This result also verifies that a single protein component, together with the RNA, is sufficient for the enzymatic telomerase activity.

30 In addition to the described TRAP assay, 5 μl of the reconstitution mixture were tested for its activity in a direct telomerase assay (Shay *et al.*, 1994). In this experiment, too, the

characteristic nucleotide ladder verifies the successful reconstitution of recombinant hTC

protein and telomerase RNA component.

In summary, it was hereby possible to demonstrate, at the functional level, that the identified,

5   and completely cloned, hTC-cDNA constitutes the catalytic subunit of human telomerase.

Reference list

Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992). Sequence identification of 2,375 human brain genes. Nature 355: 632-634.

Allsopp, R.C., Vazire, H., Pattersson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. and Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. Proc. Natl. Acad. Sci. 89, 10114-10118.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. et al. (1990). Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Ausubel, F.M., Brent, R., Kingston, R.E. Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1987). Current protocols in molecular biology. Greene Publishing Associates and Whiley-Intersciences, New York.

Blasco, M.A., Rizen, M., Greider, C.W. and Hanahan, D. (1996). Differential regulation of telomerase activity and telomerase RNA during multistage tumorigenesis. Nature Genetics 12, 200-204.

Broccoli, D., Young, J.W. and deLange, T. (1995). Telomerase activity in normal and malignant hematopoietic cells. Proc. Natl. Acad. Sci. 92, 9082-9086.

Collins, K., Kobayashi, R. and Greider, C.W. (1995). Purification of Tetrahymena telomerase and cloning of genes encoding the two protein components of the enzyme. Cell 81, 677-686.

Counter, C.M., Avilion, A.A., LeFeuvre, C.E., Stewart, N.G. Greider, C.W. Harley, C.B. and Bacchetti S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. EMBO J. 11, 1921-1929.

Counter, C.M., Gupta, J., Harley, C.B., Leber, B. and Bacchetti, S. (1995). Telomerase activity in normal leukocytes and in hematologic malignancies. Blood 85, 2315-2320.

Feng, J., Funk, W.D., Wang, S.-S., Weinrich, S.L., Avilion, A.A., Chiu, C.-P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J., Le, S., West, M.D., Harley, C.B., Andrews, W.H., Greider, C.W. and Villeponteau, B. (1995). The RNA component of human telomerase. Science 269, 1236-1241.

Gerhold, D. and Caskey, T. (1996). It's the genes! EST access to human genome content. BioEssays 18, 973-981.

Goldstein, S. (1990). Replicative senescence: The human fibroblast comes of age. Science 249, 1129-1133.

Greider, C.W. and Blackburn, E.H. (1985). Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. Cell 43, 405-413.

Greider, C.W. and Blackburn, E.H. (1987). The telomere terminal transferase of Tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. Cell 51, 887-898.

Greider, C.W. and Blackburn, E.H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. Nature 337, 331-337.

Harley, C.B., Futcher, A.B. and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature 345, 458-460.

Harrington, L., McPhail, T., Mar, V., Zhou, W., Oulton, R., Amgen EST Program, Bass, M.B., Arruda, I. and Robinson, M.O. (1997). A mammalian telomerase-associated protein. Science 275: 973-977.

Hastie, N.D., Dempster, M., Dunlop, M.G., Thompson, A.M., Green, D.K. and Allshire, R.C. (1990). Telomere reduction in human colorectal carcinoma and with ageing. Nature 346, 866-868.

Hiyama, K., Hirai, Y., Kyoizumi, S., Akiyama, M., Hiyama, E., Piatyszek, M.A., Shay, J.W., Ishioka, S. and Yamakido, M. (1995). Activation of telomerase in human lymphocytes and hematopoietic progenitor cells. J. Immunol. 155, 3711-3715.

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L.C., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994). Specific association of human telomerase activity with immortal cells and cancer. Science 266, 2011-2015.

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V. and Cech T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. Science 276: 561-567.

Lundblad, V. and Szostak, J.W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. Cell 57, 633-643.

McClintock, B. (1941). The stability of broken ends of chromosomes in *Zea mays*. Genetics 26, 234-282.

Meyne, J., Ratliff, R.L. and Moyzis, R.K. (1989). Conservation of the human telomere sequence (TTAGGG)$_n$ among vertebrates. Proc. Natl. Acad. Sci. 86, 7049-7053.

Okubo, K. Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y. and Matsubra, K. (1992). Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. Nature Genetics 2: 173-179.

Olovnikov, A.M. (1973). A theory of marginotomy. J. Theor. Biol. 41, 181-190.

Poch, O., Sauvaget, I., Delarue, M. and Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. EMBO J. 8: 3867-3874.

Pokrovskaya, I.D. and Gurevich, V.V. (1994). In vitro transcription: Preparative RNA yields in analytical scale reactions. Analytical Biochemistry 220, 420-423.

Prowse, K.R., Avilion, A.A. and Greider, C.W. (1993). Identification of a non- processive telomerase activity from mouse cells. Proc. Natl. Acad. Sci. 90, 1493-1497.

Sandell, L.L. and Zakian, V.A. (1993). Loss of a yeast telomere: Arrest, recovery and chromosome loss. Cell 75, 729-739.

Shampay, J. and Blackburn, E.H. (1988). Generation of telomere-length heterogeneity in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. 85, 534-538.

Shay, J.W., Brasiskyte, D., Ouellette, M., Piatyszek, M.A, Werbin, H., Ying, Y. and Wright, E.W. (1994). Analysis of telomerase and telomeres. Methods of Molecular Genetics 5, 263-280

Shay, J.W. (1997). Telomerae and Cancer. Ciba Foundation Meeting: Telomeres and Telomerase. London.

Singer, M.S. and Gottschling, D.E. (1994). TLC1: Template RNA Component of Saccharomyces cerevisiae Telomerase. Science 266, 404-409.

Vaziri, H., Dragowska, W., Allsopp, R.C., Thomas, T.E., Harley, C.B. and Landsdorp, P.M. (1994). Evidence for a mitotic clock in human hematopoietic stem cells: Loss of telomeric DNA with age. Proc. Natl. Acad. Sci. 91, 9857-9860.

Xiong, Y. and Eickbush, T.H. (1990). Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. 9: 3353-3362.

Yu, G.-L., Bradley, J.D., Attardi, L.D. and Blackburn, E.H. (1990). *In vivo* alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs. Nature 344, 126-132.

Zakian, V. A. (1995). Telomeres: Beginning to understand the end. Science 270, 1601-1607.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

    (i) APPLICANT:
        (A) NAME: Bayer AG
        (B) STREET: Bayerwerk
        (C) CITY: Leverkusen
        (E) COUNTRY: Germany
        (F) POSTAL CODE: D-51368
        (G) TELEPHONE: 0214-303688
        (H) TELEFAX: 0214-303482

    (ii) TITLE OF THE INVENTION: Human catalytic telomerase subunit and
      its diagnostic and therapeutic use

    (iii) NUMBER OF SEQUENCES: 7

    (iv) COMPUTER-READABLE FORM:
        (A) MEDIUM TYPE: Floppy disk
        (B) COMPUTER: IBM PC compatible
        (C) OPERATING SYSTEM: PC-DOS/MS-DOS
        (D) SOFTWARE: PatentIn Release #1.0, Version #1.30B (EPA)

(2) INFORMATION FOR SEQ ID NO: 1:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 4042 Basenpaare
        (B) TYPE: Nucleotide
        (C) STRANDEDNESS: Einzelstrang
        (D) TOPOLOGY: Linear

    (ii) ART DES MOLEKŠLS: cDNA

    (iii) HYPOTHETICAL: NO

    (iv) ANTISENSE: NO

    (vi) ORIGINAL SOURCE:
        (C) INDIVIDUAL/ISOLATE: Human

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```
GTTTCAGGCA GCGCTGCGTC CTGCTGCGCA CGTGGGAAGC CCTGGCCCCG GCCACCCCCG      60

CGATGCCGCG CGCTCCCCGC TGCCGAGCCG TGCGCTCCCT GCTGCGCAGC CACTACCGCG     120

AGGTGCTGCC GCTGGCCACG TTCGTGCGGC GCCTGGGGCC CCAGGGCTGG CGGCTGGTGC     180

AGCGCGGGGA CCCGGCGGCT TTCCGCGCGC TGGTGGCCCA GTGCCTGGTG TGCGTGCCCT     240

GGGACGCACG GCCGCCCCCC GCCGCCCCCT CCTTCCGCCA GGTGTCCTGC CTGAAGGAGC     300

TGGTGGCCCG AGTGCTGCAG AGGCTGTGCG AGCGCGGCGC GAAGAACGTG CTGGCCTTCG     360

GCTTCGCGCT GCTGGACGGG GCCCGCGGGG GCCCCCCCGA GGCCTTCACC ACCAGCGTGC     420

GCAGCTACCT GCCCAACACG GTGACCGACG CACTGCGGGG GAGCGGGGCG TGGGGGCTGC     480
```

```
TGCTGCGCCG  CGTGGGCGAC  GACGTGCTGG  TTCACCTGCT  GGCACGCTGC  GCGCTCTTTG    540

TGCTGGTGGC  TCCCAGCTGC  GCCTACCAGG  TGTGCGGGCC  GCCGCTGTAC  CAGCTCGGCG    600

CTGCCACTCA  GGCCCGGCCC  CCGCCACACG  CTAGTGGACC  CCGAAGGCGT  CTGGGATGCG    660

AACGGGCCTG  GAACCATAGC  GTCAGGGAGG  CCGGGGTCCC  CCTGGGCCTG  CCAGCCCCGG    720

GTGCGAGGAG  GCGCGGGGGC  AGTGCCAGCC  GAAGTCTGCC  GTTGCCCAAG  AGGCCCAGGC    780

GTGGCGCTGC  CCCTGAGCCG  GAGCGGACGC  CCGTTGGGCA  GGGGTCCTGG  GCCCACCCGG    840

GCAGGACGCG  TGGACCGAGT  GACCGTGGTT  TCTGTGTGGT  GTCACCTGCC  AGACCCGCCG    900

AAGAAGCCAC  CTCTTTGGAG  GGTGCGCTCT  CTGGCACGCG  CCACTCCCAC  CCATCCGTGG    960

GCCGCCAGCA  CCACGCGGGC  CCCCCATCCA  CATCGCGGCC  ACCACGTCCC  TGGGACACGC   1020

CTTGTCCCCC  GGTGTACGCC  GAGACCAAGC  ACTTCCTCTA  CTCCTCAGGC  GACAAGGAGC   1080

AGCTGCGGCC  CTCCTTCCTA  CTCAGCTCTC  TGAGGCCCAG  CCTGACTGGC  GCTCGGAGGC   1140

TCGTGGAGAC  CATCTTTCTG  GGTTCCAGGC  CCTGGATGCC  AGGGACTCCC  CGCAGGTTGC   1200

CCCGCCTGCC  CCAGCGCTAC  TGGCAAATGC  GGCCCCTGTT  TCTGGAGCTG  CTTGGGAACC   1260

ACGCGCAGTG  CCCCTACGGG  GTGCTCCTCA  AGACGCACTG  CCCGCTGCGA  GCTGCGGTCA   1320

CCCCAGCAGC  CGGTGTCTGT  GCCCGGGAGA  AGCCCCAGGG  CTCTGTGGCG  GCCCCCGAGG   1380

AGGAGGACAC  AGACCCCCGT  CGCCTGGTGC  AGCTGCTCCG  CCAGCACAGC  AGCCCCTGGC   1440

AGGTGTACGG  CTTCGTGCGG  GCCTGCCTGC  GCCGGCTGGT  GCCCCCAGGC  CTCTGGGGCT   1500

CCAGGCACAA  CGAACGCCGC  TTCCTCAGGA  ACACCAAGAA  GTTCATCTCC  CTGGGGAAGC   1560

ATGCCAAGCT  CTCGCTGCAG  GAGCTGACGT  GGAAGATGAG  CGTGCGGGAC  TGCGCTTGGC   1620

TGCGCAGGAG  CCCAGGGGTT  GGCTGTGTTC  CGGCCGCAGA  GCACCGTCTG  CGTGAGGAGA   1680

TCCTGGCCAA  GTTCCTGCAC  TGGCTGATGA  GTGTGTACGT  CGTCGAGCTG  CTCAGGTCTT   1740

TCTTTTATGT  CACGGAGACC  ACGTTTCAAA  AGAACAGGCT  CTTTTTCTAC  CGGAAGAGTG   1800

TCTGGAGCAA  GTTGCAAAGC  ATTGGAATCA  GACAGCACTT  GAAGAGGGTG  CAGCTGCGGG   1860

AGCTGTCGGA  AGCAGAGGTC  AGGCAGCATC  GGGAAGCCAG  GCCCGCCCTG  CTGACGTCCA   1920

GACTCCGCTT  CATCCCCAAG  CCTGACGGGC  TGCGGCCGAT  TGTGAACATG  GACTACGTCG   1980

TGGGAGCCAG  AACGTTCCGC  AGAGAAAAGA  GGGCCGAGCG  TCTCACCTCG  AGGGTGAAGG   2040

CACTGTTCAG  CGTGCTCAAC  TACGAGCGGG  CGCGGCGCCC  CGGCCTCCTG  GGCGCCTCTG   2100

TGCTGGGCCT  GGACGATATC  CACAGGGCCT  GGCGCACCTT  CGTGCTGCGT  GTGCGGGCCC   2160

AGGACCCGCC  GCCTGAGCTG  TACTTTGTCA  AGGTGGATGT  GACGGGCGCG  TACGACACCA   2220

TCCCCCAGGA  CAGGCTCACG  GAGGTCATCG  CCAGCATCAT  CAAACCCCAG  AACACGTACT   2280

GCGTGCGTCG  GTATGCCGTG  GTCCAGAAGG  CCGCCCATGG  GCACGTCCGC  AAGGCCTTCA   2340
```

```
AGAGCCACGT CTCTACCTTG ACAGACCTCC AGCCGTACAT GCGACAGTTC GTGGCTCACC    2400

TGCAGGAGAC CAGCCCGCTG AGGGATGCCG TCGTCATCGA GCAGAGCTCC TCCCTGAATG    2460

AGGCCAGCAG TGGCCTCTTC GACGTCTTCC TACGCTTCAT GTGCCACCAC GCCGTGCGCA    2520

TCAGGGGCAA GTCCTACGTC CAGTGCCAGG GGATCCCGCA GGGCTCCATC CTCTCCACGC    2580

TGCTCTGCAG CCTGTGCTAC GGCGACATGG AGAACAAGCT GTTTGCGGGG ATTCGGCGGG    2640

ACGGGCTGCT CCTGCGTTTG GTGGATGATT TCTTGTTGGT GACACCTCAC CTCACCCACG    2700

CGAAAACCTT CCTCAGGACC CTGGTCCGAG GTGTCCCTGA GTATGGCTGC GTGGTGAACT    2760

TGCGGAAGAC AGTGGTGAAC TTCCCTGTAG AAGACGAGGC CCTGGGTGGC ACGGCTTTTG    2820

TTCAGATGCC GGCCCACGGC CTATTCCCCT GGTGCGGCCT GCTGCTGGAT ACCCGGACCC    2880

TGGAGGTGCA GAGCGACTAC TCCAGCTATG CCCGGACCTC CATCAGAGCC AGTCTCACCT    2940

TCAACCGCGG CTTCAAGGCT GGGAGCAACA TGCGTCGCAA ACTCTTTGGG GTCTTGCGGC    3000

TGAAGTGTCA CAGCCTGTTT CTGGATATGC AGGTGAACAG CCTCCAGACG GTGTGCACCA    3060

ACATCTACAA GATCCTCCTG CTGCAGGCGT ACAGGTTTCA CGCATGTGTG CTGCAGCTCC    3120

CATTTCATCA GCAAGTTTGG AAGAACCCCA CATTTTTCCT GCGCGTCATC TCTGACACGG    3180

CCTCCCTCTG CTACTCCATC CTGAAAGCCA AGAACGCAGG GATGTCGCTG GGGGCCAAGG    3240

GCGCCGCCGG CCCTCTGCCC TCCGAGGCCG TGCAGTGGCT GTGCCACCAA GCATTCCTGC    3300

TCAAGCTGAC TCGACACCGT GTCACCTACG TGCCACTCCT GGGGTCACTC AGGACAGCCC    3360

AGACGCAGCT GAGTCGGAAG CTCCCGGGGA CGACGCTGAC TGCCCTGGAG GCCGCAGCCA    3420

ACCCGGCACT GCCCTCAGAC TTCAAGACCA TCCTGGACTG ATGGCCACCC GCCCACAGCC    3480

AGGCCGAGAG CAGACACCAG CAGCCCTGTC ACGCCGGGCT CTACGTCCCA GGGAGGGAGG    3540

GGCGGCCCAC ACCCAGGCCC GCACCGCTGG GAGTCTGAGG CCTGAGTGAG TGTTTGGCCG    3600

AGGCCTGCAT GTCCGGCTGA AGGCTGAGTG TCCGGCTGAG GCCTGAGCGA GTGTCCAGCC    3660

AAGGGCTGAG TGTCCAGCAC ACCTGCCGTC TTCACTTCCC CACAGGCTGG CGCTCGGCTC    3720

CACCCCAGGG CCAGCTTTTC CTCACCAGGA GCCCGGCTTC CACTCCCCAC ATAGGAATAG    3780

TCCATCCCCA GATTCGCCAT TGTTCACCCC TCGCCCTGCC CTCCTTTGCC TTCCACCCCC    3840

ACCATCCAGG TGGAGACCCT GAGAAGGACC CTGGGAGCTC TGGGAATTTG GAGTGACCAA    3900

AGGTGTGCCC TGTACACAGG CGAGGACCCT GCACCTGGAT GGGGGTCCCT GTGGGTCAAA    3960

TTGGGGGGAG GTGCTGTGGG AGTAAAATAC TGAATATATG AGTTTTTCAG TTTTGAAAAA    4020

                                                                   4042

AAAAAAAAAA AAAAAAAAAA AA
```

(2) INFORMATION FOR SEQ ID NO: 2:

  (i) SEQUENCE CHARACTERISTICS:

- 48 -

```
        (A) LENGTH: 1132 amino acids
        (B) TYPE: Amino acid
        (C) STRANDEDNESS: Single
        (D) TOPOLOGY: Linear

 (ii) ART DES MOLEKŠLS: Protein

(iii) HYPOTHETICAL: NO

 (iv) ANTISENSE: NO

 (vi) ORIGINAL SOURCE:
        (C) INDIVIDUAL/ISOLATE: Human


 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Pro Arg Ala Pro Arg Cys Arg Ala Val Arg Ser Leu Leu Arg Ser
1               5               10              15

His Tyr Arg Glu Val Leu Pro Leu Ala Thr Phe Val Arg Arg Leu Gly
            20              25              30

Pro Gln Gly Trp Arg Leu Val Gln Arg Gly Asp Pro Ala Ala Phe Arg
        35              40              45

Ala Leu Val Ala Gln Cys Leu Val Cys Val Pro Trp Asp Ala Arg Pro
    50              55              60

Pro Pro Ala Ala Pro Ser Phe Arg Gln Val Ser Cys Leu Lys Glu Leu
65              70              75              80

Val Ala Arg Val Leu Gln Arg Leu Cys Glu Arg Gly Ala Lys Asn Val
            85              90              95

Leu Ala Phe Gly Phe Ala Leu Leu Asp Gly Ala Arg Gly Gly Pro Pro
            100             105             110

Glu Ala Phe Thr Thr Ser Val Arg Ser Tyr Leu Pro Asn Thr Val Thr
            115             120             125

Asp Ala Leu Arg Gly Ser Gly Ala Trp Gly Leu Leu Leu Arg Arg Val
    130             135             140

Gly Asp Asp Val Leu Val His Leu Leu Ala Arg Cys Ala Leu Phe Val
145             150             155             160

Leu Val Ala Pro Ser Cys Ala Tyr Gln Val Cys Gly Pro Pro Leu Tyr
            165             170             175

Gln Leu Gly Ala Ala Thr Gln Ala Arg Pro Pro Pro His Ala Ser Gly
            180             185             190

Pro Arg Arg Arg Leu Gly Cys Glu Arg Ala Trp Asn His Ser Val Arg
        195             200             205

Glu Ala Gly Val Pro Leu Gly Leu Pro Ala Pro Gly Ala Arg Arg Arg
    210             215             220

Gly Gly Ser Ala Ser Arg Ser Leu Pro Leu Pro Lys Arg Pro Arg Arg
225             230             235             240
```

Gly Ala Ala Pro Glu Pro Glu Arg Thr Pro Val Gly Gln Gly Ser Trp
           245   250   255

Ala His Pro Gly Arg Thr Arg Gly Pro Ser Asp Arg Gly Phe Cys Val
           260   265   270

Val Ser Pro Ala Arg Pro Ala Glu Glu Ala Thr Ser Leu Glu Gly Ala
       275   280   285

Leu Ser Gly Thr Arg His Ser His Pro Ser Val Gly Arg Gln His His
       290   295   300

Ala Gly Pro Pro Ser Thr Ser Arg Pro Pro Arg Pro Trp Asp Thr Pro
305   310   315   320

Cys Pro Pro Val Tyr Ala Glu Thr Lys His Phe Leu Tyr Ser Ser Gly
         325   330   335

Asp Lys Glu Gln Leu Arg Pro Ser Phe Leu Leu Ser Ser Leu Arg Pro
       340   345   350

Ser Leu Thr Gly Ala Arg Arg Leu Val Glu Thr Ile Phe Leu Gly Ser
       355   360   365

Arg Pro Trp Met Pro Gly Thr Pro Arg Arg Leu Pro Arg Leu Pro Gln
   370   375   380

Arg Tyr Trp Gln Met Arg Pro Leu Phe Leu Glu Leu Leu Gly Asn His
385   390   395   400

Ala Gln Cys Pro Tyr Gly Val Leu Leu Lys Thr His Cys Pro Leu Arg
         405   410   415

Ala Ala Val Thr Pro Ala Ala Gly Val Cys Ala Arg Glu Lys Pro Gln
       420   425   430

Gly Ser Val Ala Ala Pro Glu Glu Glu Asp Thr Asp Pro Arg Arg Leu
       435   440   445

Val Gln Leu Leu Arg Gln His Ser Ser Pro Trp Gln Val Tyr Gly Phe
   450   455   460

Val Arg Ala Cys Leu Arg Arg Leu Val Pro Pro Gly Leu Trp Gly Ser
465   470   475   480

Arg His Asn Glu Arg Arg Phe Leu Arg Asn Thr Lys Lys Phe Ile Ser
           485   490   495

Leu Gly Lys His Ala Lys Leu Ser Leu Gln Glu Leu Thr Trp Lys Met
           500   505   510

Ser Val Arg Asp Cys Ala Trp Leu Arg Arg Ser Pro Gly Val Gly Cys
       515   520   525

Val Pro Ala Ala Glu His Arg Leu Arg Glu Glu Ile Leu Ala Lys Phe
       530   535   540

Leu His Trp Leu Met Ser Val Tyr Val Val Glu Leu Leu Arg Ser Phe
545   550   555   560

Phe Tyr Val Thr Glu Thr Thr Phe Gln Lys Asn Arg Leu Phe Phe Tyr
       565   570   575

Arg Lys Ser Val Trp Ser Lys Leu Gln Ser Ile Gly Ile Arg Gln His
            580   585   590

Leu Lys Arg Val Gln Leu Arg Glu Leu Ser Glu Ala Glu Val Arg Gln
            595   600   605

His Arg Glu Ala Arg Pro Ala Leu Leu Thr Ser Arg Leu Arg Phe Ile
    610   615   620

Pro Lys Pro Asp Gly Leu Arg Pro Ile Val Asn Met Asp Tyr Val Val
625   630   635   640

Gly Ala Arg Thr Phe Arg Arg Glu Lys Arg Ala Glu Arg Leu Thr Ser
            645   650   655

Arg Val Lys Ala Leu Phe Ser Val Leu Asn Tyr Glu Arg Ala Arg Arg
            660   665   670

Pro Gly Leu Leu Gly Ala Ser Val Leu Gly Leu Asp Asp Ile His Arg
            675   680   685

Ala Trp Arg Thr Phe Val Leu Arg Val Arg Ala Gln Asp Pro Pro Pro
    690   695   700

Glu Leu Tyr Phe Val Lys Val Asp Val Thr Gly Ala Tyr Asp Thr Ile
705   710   715   720

Pro Gln Asp Arg Leu Thr Glu Val Ile Ala Ser Ile Ile Lys Pro Gln
            725   730   735

Asn Thr Tyr Cys Val Arg Arg Tyr Ala Val Val Gln Lys Ala Ala His
            740   745   750

Gly His Val Arg Lys Ala Phe Lys Ser His Val Ser Thr Leu Thr Asp
            755   760   765

Leu Gln Pro Tyr Met Arg Gln Phe Val Ala His Leu Gln Glu Thr Ser
    770   775   780

Pro Leu Arg Asp Ala Val Val Ile Glu Gln Ser Ser Ser Leu Asn Glu
785   790   795   800

Ala Ser Ser Gly Leu Phe Asp Val Phe Leu Arg Phe Met Cys His His
            805   810   815

Ala Val Arg Ile Arg Gly Lys Ser Tyr Val Gln Cys Gln Gly Ile Pro
            820   825   830

Gln Gly Ser Ile Leu Ser Thr Leu Leu Cys Ser Leu Cys Tyr Gly Asp
            835   840   845

Met Glu Asn Lys Leu Phe Ala Gly Ile Arg Arg Asp Gly Leu Leu Leu
    850   855   860

Arg Leu Val Asp Asp Phe Leu Leu Val Thr Pro His Leu Thr His Ala
865   870   875   880

Lys Thr Phe Leu Arg Thr Leu Val Arg Gly Val Pro Glu Tyr Gly Cys
            885   890   895

Val Val Asn Leu Arg Lys Thr Val Val Asn Phe Pro Val Glu Asp Glu

```
                    900               905               910
Ala Leu Gly Gly Thr Ala Phe Val Gln Met Pro Ala His Gly Leu Phe
        915             920             925
Pro Trp Cys Gly Leu Leu Leu Asp Thr Arg Thr Leu Glu Val Gln Ser
    930             935             940
Asp Tyr Ser Ser Tyr Ala Arg Thr Ser Ile Arg Ala Ser Leu Thr Phe
945             950             955             960
Asn Arg Gly Phe Lys Ala Gly Arg Asn Met Arg Arg Lys Leu Phe Gly
                965             970             975
Val Leu Arg Leu Lys Cys His Ser Leu Phe Leu Asp Leu Gln Val Asn
        980             985             990
Ser Leu Gln Thr Val Cys Thr Asn Ile Tyr Lys Ile Leu Leu Leu Gln
        995             1000            1005
Ala Tyr Arg Phe His Ala Cys Val Leu Gln Leu Pro Phe His Gln Gln
1010            1015            1020
Val Trp Lys Asn Pro Thr Phe Phe Leu Arg Val Ile Ser Asp Thr Ala
1025            1030            1035            1040
Ser Leu Cys Tyr Ser Ile Leu Lys Ala Lys Asn Ala Gly Met Ser Leu
            1045            1050            1055
Gly Ala Lys Gly Ala Ala Gly Pro Leu Pro Ser Glu Ala Val Gln Trp
            1060            1065            1070
Leu Cys His Gln Ala Phe Leu Leu Lys Leu Thr Arg His Arg Val Thr
            1075            1080            1085
Tyr Val Pro Leu Leu Gly Ser Leu Arg Thr Ala Gln Thr Gln Leu Ser
            1090            1095            1100
Arg Lys Leu Pro Gly Thr Thr Leu Thr Ala Leu Glu Ala Ala Ala Asn
1105            1110            1115            1120
Pro Ala Leu Pro Ser Asp Phe Lys Thr Ile Leu Asp
            1125            1130
```

(2) INFORMATION FOR SEQ ID NO: 3:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1153 base pairs
        (B) TYPE: Nucleotide
        (C) STRANDEDNESS: Single
        (D) TOPOLOGY: Linear

    (ii) MOLECULE TYPE: cDNA

   (iii) HYPOTHETICAL: NO

    (iv) ANTISENSE: NO

    (vi) ORIGINAL SOURCE:
        (C) INDIVIDUAL/ISOLATE: Human

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
GTGCCTGCAG AGACCGGTCT GGTGCACTCT GATTCTCCAC TTGCCTGTTG CATGTCCTCG      60

TTCCCTTGTT TCTCACCACC TCTTGGGTTG CCATGTGCGT TTCCTGCCGA GTGTGTGTTG     120

ATCCTCTCGT TGCCTCCTGG TCACTGGGCA TTTGCTTTTA TTTCTCTTTG CTTAGTGTTA     180

CCCCCTGATC TTTTTATTGT CGTTGTTTGC TTTTGTTTAT TGAGACAGTC TCACTCTGTC     240

ACCCAGGCTG GAGTGTAATG GCACAATCTC GGCTCACTGC AACCTCTGCC TCCTCGGTTC     300

AAGCAGTTCT CATTCCTCAA CCTCATGAGT AGCTGGGATT ACAGGCGCCC ACCACCACGC     360

CTGGCTAATT TTTGTATTTT TAGTAGAGAT AGGCTTTCAC CATGTTGGCC AGGCTGGTCT     420

CAAACTCCTG ACCTCAAGTG ATCTGCCCGC CTTGGCCTCC CACAGTGCTG GGATTACAGG     480

TGCAAGCCAC CGTGCCCGGC ATACCTTGAT CTTTTAAAAT GAAGTCTGAA ACATTGCTAC     540

CCTTGTCCTG AGCAATAAGA CCCTTACTGT ATTTTAGCTC TGGCCACCCC CCAGCCTGTG     600

TGCTGTTTTC CCTGCTGACT TAGTTCTATC TCAGGCATCT TGACACCCCC ACAAGCTAAG     660

CATTATTAAT ATTGTTTTCC GTGTTGAGTG TTTCTTTAGC TTTGCCCCCG CCCTGCTTTT     720

CCTCCTTTGT TCCCCGTCTG TCTTCTGTCT CAGGCCCGCC GTCTGGGGTC CCCTTCCTTG     780

TCCTTTGCGT GGTTCTTCTG TCTTGTTATT GCTGGTAAAC CCCAGCTTTA CCTGTGCTGG     840

CCTCCATGGC ATCTAGCGAC GTCCGGGGAC CTCTGCTTAT GATGCACAGA TGAAGATGTG     900

GAGACTCACG AGGAGGGCGG TCATCTTGGC CCGTGAGTGT CTGGAGCACC ACGTGGCCAG     960

CGTTCCTTAG CCAGGGTTGG CTGTGTTCCG GCCGCAGAGC ACCGTCTGCG TGAGGAGATC    1020

CTGGCCAAGT TCCTGCACTG GCTGATGAGT GTGTACGTCG TCGAGCTGCT CAGGTCTTTC    1080

TTTTATGTCA CGGAGACCAC GTTTCAAAAG AACAGGCTCT TTTTCTACCG GAAGAGTGTC    1140

TGGAGCAAGT TGC                                                      1153
```

(2) INFORMATION FOR SEQ ID NO: 4:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 412 base pairs
        (B) TYPE: Nucleotide
        (C) STRANDEDNESS: Single
        (D) TOPOLOGY: Linear

    (ii) MOLECULE TYPE: cDNA

   (iii) HYPOTHETICAL: NO

    (iv) ANTISENSE: NO

    (vi) ORIGINAL SOURCE:
        (C) INDIVIDUAL/ISOLATE: Human

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

```
CAGAGCCCTG GTCCTCCTGT CTCCATCGTC ACGTGGGCAC ACGTGGCTTT TCGCTCAGGA      60

CGTCGAGTGG ACACGGTGAT CTCTGCCTCT GCTCTCCCTC CTGTCCAGTT TGCATAAACT     120

TACGAGGTTC ACCTTCACGT TTTGATGGAC ACGCGGTTTC CAGGCACCGA GGCCAGAGCA     180

GTGAACAGAG GAGGCTGGGC GCGGCAGTGG AGCCGGGTTG CCGGCAATGG GGAGAAGTGT     240

CTGGAAGCAC AGACGCTCTG GCGAGGGTGC CTGCAGAGAC CCGCCTGGTG CACTCTGATT     300

CTCCACTTGC CTGTTGCATG TCCTCGTTCC CTTGTTTCTC ACCACCTCTT GGGTTGCCAT     360

GTGCGTTTCC TGCCGAGTGT GTGTTGATCC TCTCGTTGCC TCCTGGTCAC TG            412
```

(2) INFORMATION FOR SEQ ID NO: 5:

     (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 1012 base pairs
        (B) TYPE: Nucleotide
        (C) STRANDEDNESS: Single
        (D) TOPOLOGY: Linear

    (ii) MOLCULE TYPE: cDNA

    (iii) HYPOTHETICAL: NO

    (iv) ANTISENSE: NO

    (vi) ORIGINAL SOURCE:
        (C) INDIVIDUAL/ISOLATE: Human

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```
GGGGTCCTGG GCCCACCCGG GCAGGACGCG TGGACCGAGT GACCGTGGTT TCTGTGTGGT      60

GTCACCTGCC AGACCCGCCG AAGAAGCCAC CTCTTTGGAG GGTGCGCTCT CTGGCACGCG     120

CCACTCCCAC CCATCCGTGG GCCGCCAGCA CCACGCGGGC CCCCCATCCA CATCGCGGCC     180

ACCACGTCCC TGGGACACGC CTTGTCCCCC GGTGTACGCC GAGACCAAGC ACTTCCTCTA     240

CTCCTCAGGC GACAAGGAGC AGCTGCGGCC CTCCTTCCTA CTCAGCTCTC TGAGGCCCAG     300

CCTGACTGGC GCTCGGAGGC TCGTGGAGAC CATCTTTCTG GGTTCCAGGC CCTGGATGCC     360

AGGGACTCCC CGCAGGTTGC CCCGCCTGCC CCAGCGCTAC TGGCAAATGC GGCCCCTGTT     420

TCTGGAGCTG CTTGGGAACC ACGCGCAGTG CCCCTACGGG GTGCTCCTCA AGACGCACTG     480

CCCGCTGCGA GCTGCGGTCA CCCCAGCAGC CGGTGTCTGT GCCCGGGAGA AGCCCCAGGG     540

CTCTGTGGCG GCCCCCGAGG AGGAGGACAC AGACCCCCGT CGCCTGGTGC AGCTGCTCCG     600

CCAGCACAGC AGCCCCTGGC AGGTGTACGG CTTCGTGCGG GCCTGCCTGC GCCGGCTGGT     660

GCCCCCAGGC CTCTGGGGCT CCAGGCACAA CGAACGCCGC TTCCTCAGGA ACACCAAGAA     720

GTTCATCTCC CTGGGGAAGC ATGCCAAGCT CTCGCTGCAG GAGCTGACGT GGAAGATGAG     780

CGTGCGGGAC TGCGCTTGGC TGCGCAGGAG CCCAGGTGAG GAGGTGGTGG CCGTCGAGGG     840
```

```
CCCAGGCCCC AGAGCTGAAT GCAGTAGGGG CTCAGAAAAG GGGGCAGGCA GAGCCCTGGT        900

CCTCCTGTCT CCATCGTCAC GTGGGCACAC GTGGCTTTTC GCTCAGGACG TCGAGTGGAC        960

ACGGTGATCT CTGCCTCTGC TCTCCCTCCT GTCCAGTTTG CATAAACTTA CG              1012
```

(2) INFORMATION FOR SEQ ID NO: 6:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 3972 base pairs
        (B) TYPE: Nucleotide
        (C) STRANDEDNESS: Single
        (D) TOPOLOGY: Linear

    (ii) MOLECULE TYPE: cDNA

    (iii) HYPOTHETICAL: NO

    (iv) ANTISENSE: NO

    (vi) ORIGINAL SOURCE:
        (C) INDIVIDUAL/ISOLATE: Human

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
GAATTCGCGG CCGCGTCGAC GTTTCAGGCA GCGCTGCGTC CTGCTGCGCA CGTGGGAAGC         60

CCTGGCCCCG GCCACCCCCG CGATGCCGCG CGCTCCCCGC TGCCGAGCCG TGCGCTCCCT        120

GCTGCGCAGC CACTACCGCG AGGTGCTGCC GCTGGCCACG TTCGTGCGGC GCCTGGGGCC        180

CCAGGGCTGG CGGCTGGTGC AGCGCGGGGA CCCGGCGGCT TTCCGCGCGC TGGTGGCCCA        240

GTGCCTGGTG TGCGTGCCCT GGGACGCACG GCCGCCCCCC GCCGCCCCCT CCTTCCGCCA        300

GGTGTCCTGC CTGAAGGAGC TGGTGGCCCG AGTGCTGCAG AGGCTGTGCG AGCGCGGCGC        360

GAAGAACGTG CTGGCCTTCG GCTTCGCGCT GCTGGACGGG GCCCGCGGGG GCCCCCCCGA        420

GGCCTTCACC ACCAGCGTGC GCAGCTACCT GCCCAACACG GTGACCGACG CACTGCGGGG        480

GAGCGGGGCG TGGGGGCTGC TGCTGCGCCG CGTGGGCGAC GACGTGCTGG TTCACCTGCT        540

GGCACGCTGC GCGCTCTTTG TGCTGGTGGC TCCCAGCTGC GCCTACCAGG TGTGCGGGCC        600

GCCGCTGTAC CAGCTCGGCG CTGCCACTCA GGCCCGGCCC CCGCCACACG CTAGTGGACC        660

CCGAAGGCGT CTGGGATGCG AACGGGCCTG GAACCATAGC GTCAGGGAGG CCGGGGTCCC        720

CCTGGGCCTG CCAGCCCCGG GTGCGAGGAG GCGCGGGGGC AGTGCCAGCC GAAGTCTGCC        780

GTTGCCCAAG AGGCCCAGGC GTGGCGCTGC CCCTGAGCCG GAGCGGACGC CCGTTGGGCA        840

GGGGTCCTGG GCCCACCCGG GCAGGACGCG TGGACCGAGT GACCGTGGTT TCTGTGTGGT        900

GTCACCTGCC AGACCCGCCG AAGAAGCCAC CTCTTTGGAG GGTGCGCTCT CTGGCACGCG        960

CCACTCCCAC CCATCCGTGG CCGCCAGCA CCACGCGGGC CCCCCATCCA CATCGCGGCC       1020
```

```
ACCACGTCCC TGGGACACGC CTTGTCCCCC GGTGTACGCC GAGACCAAGC ACTTCCTCTA    1080

CTCCTCAGGC GACAAGGAGC AGCTGCGGCC CTCCTTCCTA CTCAGCTCTC TGAGGCCCAG    1140

CCTGACTGGC GCTCGGAGGC TCGTGGAGAC CATCTTTCTG GGTTCCAGGC CCTGGATGCC    1200

AGGGACTCCC CGCAGGTTGC CCCGCCTGCC CCAGCGCTAC TGGCAAATGC GGCCCCTGTT    1260

TCTGGAGCTG CTTGGGAACC ACGCGCAGTG CCCCTACGGG GTGCTCCTCA AGACGCACTG    1320

CCCGCTGCGA GCTGCGGTCA CCCCAGCAGC CGGTGTCTGT GCCCGGGAGA AGCCCCAGGG    1380

CTCTGTGGCG GCCCCCGAGG AGGAGGACAC AGACCCCGT CGCCTGGTGC AGCTGCTCCG      1440

CCAGCACAGC AGCCCCTGGC AGGTGTACGG CTTCGTGCGG GCCTGCCTGC GCCGGCTGGT    1500

GCCCCCAGGC CTCTGGGGCT CCAGGCACAA CGAACGCCGC TTCCTCAGGA ACACCAAGAA    1560

GTTCATCTCC CTGGGGAAGC ATGCCAAGCT CTCGCTGCAG GAGCTGACGT GGAAGATGAG    1620

CGTGCGGGAC TGCGCTTGGC TGCGCAGGAG CCCAGGTGAG GAGGTGGTGG CCGTCGAGGG    1680

CCCAGGCCCC AGAGCTGAAT GCAGTAGGGG CCCAGAAAAG GGGGCAGGCA GAGCCCTGGT    1740

CCTCCTGTCT CCATCGTCAC GTGGGCACAC GCGGCTTTTC GCTCAGGACG TCGAGTGGAC    1800

ACGGTGATCT CTGCCTCTGC TCTCCCTCCT GTCCAGTTTG CATAAACTTA CGAGGTTCAC    1860

CTTCACGTTT TGATGGACAC GCGGTTTCCA GGCGCCGAGG CCAGAGCAGT GAACAGAGGA    1920

GGCTGGGCGC GGCAGTGGAG CCGGGTTGCC GGCAATGGGG AGAAGTGTCT GGAAGCACAG    1980

ACGCTCTGGC GAGGGTGCCT GCAGGGGTTG GCTGTGTTCC GGCCGCAGAG CACCGTCTGC    2040

GTGAGGAGAT CCTGGCCAAG TTCCTGCACT GGCTGATGAG TGTGTACGTC GTCGAGCTGC    2100

TCAGGTCTTT CTTTTATGTC ACGGAGACCA CGTTTCAAAA GAACAGGCTC TTTTTCTACC    2160

GGAAGAGTGT CTGGAGCAAG TTGCAAAGCA TTGGAATCAG ACAGCACTTG AAGAGGGTGC    2220

AGCTGCGGGA GCTGTCGGAA GCAGAGGTCA GGCAGCATCG GGAAGCCAGG CCCGCCCTGC    2280

TGACGTCCAG ACTCCGCTTC ATCCCCAAGC CTGACGGGCT GCGGCCGATT GTGAACATGG    2340

ACTACGTCGT GGGAGCCAGA ACGTTCCGCA GAGAAAAGAG GGTGGCTGTG CTTTGGTTTA    2400

ACTTCCTTTT TAAACAGAAG TGCGTTTGAG CCCCACATTT GGTATCAGCT TAGATGAAGG    2460

GCCCGGAGGA GGGGCCACGG GACACAGCCA GGGCCATGGC ACGGCGCCAA CCCATTTGTG    2520

CGCACGGTGA GGTGGCCGAG GTGCCGGTGC CTCCAGAAAA GCAGCGTGGG GGTGTAGGGG    2580

GAGCTCCTGG GGCAGGGACA GGCTCTGAGG ACCACAAGAA GCAGCTGGGC CAGGGCCTGG    2640

ATGCAGCACG GCCCGAGCGG GTGGGGGCCC ACCACGCCAT TCTGGTCAAA GGTGTTGTAG    2700

TCGTAATAGC CGGCCCAGGC GCTCTGAACC TTCAGAGTCT CAAAAGCTGG GACCCTCAGG    2760

GCCAAATGGG GCCACACCTT GTCCTGGAAG AAATCATGGT CCACTTCCAG GTTCGCCGGG    2820

TCCGGTTCTT CCTGCTCAGT GGGGCTACGA CCACCTAGGT AGTTGCTACC TAATCCTTCC    2880
```

```
CGGCGAAAAT AGGCTCCACT GGTGTCTGCA ACAAGCGGAG TCTCTAGGCC TGGTCCCTGG    2940

GGGCAGTGCC ACACATACAG ATACCTTTTC CTCGGCTCCA CAGGTAGCTT GGTGCCCTGC    3000

AGGGTGCCAG GCGGCCCCTC TCCAACACCA GCCAGTGCTG CGATTTGCGC AGACCAGGCT    3060

CCGGCTGCGT TGATCACAAT GGCGCATTCC ACAGGCTGGT ACTCCAGGCT GCGGTCCATC    3120

TTCACATGGA CTTCATGGAT CCTTTTCAAG ACCACCGCTT TGTCATCTGT GGTCAACATG    3180

CGTTGAGATG AAGAGACAAA ACGTGTCACC TCTCCCTGGC AGAAAAGGAC TCCCAAGGAC    3240

TGGACCTTTC GCCGAAGCCC CTGGAGCAGA CACCAGGGGT CAAACCAACC TTCGTCCTCC    3300

ATCCCATAAG ACGCCAAAGC CACTCCCTCT GTGTTTATCC AGGGAAACTT GTTCCGAAGC    3360

TGATCAGGAG ACATCAGAGA AACTTTGGCT CCCTCCTGCC TCTGCACTTT CACGTTGCTC    3420

TCCATGGCTG CAGCATCCTT TTCTGAAGCC AGCAAGAGGT AGCCCGAGGG GTTGAACCGG    3480

AGGTCCAGGG GAGGAGCATC GACTACGGCC AGGTACTCAT TGATGTTCCG TAGAAAGCTG    3540

GCTGAAAAGA GGGAGAGCTG GATGTTCTCA GGCAATGAGA ACTGCTGACA AATCCCACCT    3600

ACTGAGAGCC CAGTGGAGGC CTGTGAATAC GTGTGGTCCC GTTCCACCAC TAGCACTCGA    3660

ATAGCACCTC GTCTGCTCTC CAGCTTCTTC AGCCAATAGG CCACAGACAA GCCAAGCACC    3720

CCACCTCCCA CGATCACCAC ATCCGAGTGC TCGGGAGGCA GGTGGCTGGT GTCTTGCAGT    3780

AGATCACAGG ACCTTCCAGG CAGGATCGAC TTGATCTTCT TCTTAATCTC AGACACCTTT    3840

CCATCCCAGT CCAGAGAAAA GCCTCCTCTG CGCGTGCCTG GCCTCCGGGT CAAGAGGCCC    3900

CGGCCCATGC CGTGCGGCAG AACCCTCCGA ATCATAGCCC CTCTGAGCCC GGGTCGACGC    3960

GGCCGCGAAT TC                                                        3972
```

(2) INFORMATION FOR SEQ ID NO: 7:

    (i) SEQUENCE CHARACTERISTICS:
        (A) LENGTH: 2089 base pairs
        (B) TYPE: Nucleotide
        (C) STRANDEDNESS: Single
        (D) TOPOLOGY: Linear

    (ii) MOLECULE TYPE: cDNA

    (v) FRAGMENT TYPE: Linear

    (vi) ORIGINAL SOURCE:
        (C) INDIVIDUAL/ISOLATE: Human

    (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```
CCGGAAGAGT GTCTGGAGCA AGTTGCAAAG CATTGGAATC AGACAGCACT TGAAGAGGGT      60

GCAGCTGCGG GAGCTGTCGG AAGCAGAGGT CAGGCAGCAT CGGGAAGCCA GGCCCGCCCT     120

GCTGACGTCC AGACTCCGCT TCATCCCCAA GCCTGACGGG CTGCGGCCCA TTGTGAACAT     180
```

```
GGACTACGTC GTGGGAGCCA GAACGTTCCG CAGAGAAAAG AGGGCCGAGC GTCTCACCTC      240

GAGGGTGAAG GCACTGTTCA GCGTGCTCAA CTACGAGCGG GCGCGGCGCC CCGGCCTCCT      300

GGGCGCCTCT GTGCTGGGCC TGGACGATAT CCACAGGGCC TGGCGCACCT TCGTGCTGCG      360

TGTGCGGGCC CAGGACCCGC CGCCTGAGCT GTACTTTGTC AAGGTGGATG TGACGGGCGC      420

GTACGACACC ATCCCCCAGG ACAGGTCAC GGAGGTCATC GCCAGCATCA TCAAACCCCA       480

GAACACGTAC TGCGTGCGTC GGTATGCCGT GGTCCAGAAG GCCGCCCATG GGCACGTCCG      540

CAAGGCCTTC AAGAGCCACG TCTCTACCTT GACAGACCTC CAGCCGTACA TGCGACAGTT      600

CGTGGCTCAC CTGCAGGAGA CCAGCCCGCT GAGGGGTGCC GTCGTCATCG AGCAGAGCTC      660

CTCCCTGAAT GAGGCCAGCA GTGGCCTCTT CGACGTCTTC CTACGCTTCA TGTGCCACCA      720

CGCCGTGCGC ATCAGGGGCA AGTCCTACGT CCAGTGCCAG GGGATCCCGC AGGGCTCCAT      780

CCTCTCCACG CTGCTCTGCA GCCTGTGCTA CGGCGACATG GAGAACAAGC TGTTTGCGGG      840

GATTCGGCGG GACGGGCTGC TCCTGCGTTT GGTGGATGAT TTCTTGTTGG TGACACCTCA      900

CCTCACCCAC GCGAAAACCT TCCTCAGGAC CCTGGTCCGA GGTGTCCCTG AGTATGGCTG      960

CGTGGTGAAC TTGCGGAAGA CAGTGGTGAA CTTCCCTGTA GAAGACGAGG CCCTGGGTGG     1020

CACGGCTTTT GTTCAGATGC CGGCCCACGG CCTATTCCCC TGGTGCGGCC TGCTGCTGGA     1080

TACCCGGACC CTGGAGGTGC AGAGCGACTA CTCCAGCTAT GCCCGGACCT CCATCAGAGC     1140

CAGTCTCACC TTCAACCGCG GCTTCAAGGC TGGGAGGAAC ATGCGTCGCA AACTCTTTGG     1200

GGTCTTGCGG CTGAAGTGTC ACAGCCTGTT TCTGGATTTG CAGGTGAACA GCCTCCAGAC     1260

GGTGTGCACC AACATCTACA AGATCCTCCT GCTGCAGGCG TACAGGTTTC ACGCATGCGT     1320

GCTGCAGCTC CCATTTCATC AGCAAGTTTG GAAGAACCGC ACATTTTCC TGCGCGTCAT      1380

CTCTGACACG GCCTCCCTCT GCTACTCCAT CCTGAAAGCC AAGAACGCAG GTATGTGCAG     1440

GTGCCTGGCC TCAGTGGCAG CAGTGCCTGC CTGCTGGTGT TAGTGTGTCA GGAGACTGAG     1500

TGAATCTGGG CTTAGGAAGT TCTTACCCCT TTTCGCATCA GGAAGTGGTT TAACCCAACC     1560

ACTGTCAGGC TCGTCTGCCC GCCCTCTCGT GGGGTGAGCA GAGCACCTGA TGGAAGGGAC     1620

AGGAGCTGTC TGGGAGCTGC CATCCTTCCC ACCTTGCTCT GCCTGGGGAA GCGCTGGGGG     1680

GCCTGGTCTC TCCTGTTTGC CCCATGGTGG GATTTGGGGG GCCTGGCCTC TCCTGTTTGC     1740

CCTGTGGTGG GATTGGGCTG TCTCCCGTCC ATGGCACTTA GGGCCCTTGT GCAAACCCAG     1800

GCCAAGGGCT TAGGAGGAGG CCAGGCCCAG GCTACCCCAC CCCTCTCAGG AGCAGAGGCC     1860

GCGTATCACC ACGACAGAGC CCCGCGCCGT CCTCTGCTTC CCAGTCACCG TCCTCTGCCC     1920

CTGGACACTT TGTCCAGCAT CAGGGAGGTT TCTGATCCGT CTGAAATTCA AGCCATGTCG     1980

AACCTGCGGT CCTGAGCTTA ACAGCTTCTA CTTTCTGTTC TTTCTGTGTT GTGGAGACCC     2040
```

TGAGAAGGAC CCTGGGAGCT CTGGGAATTT GGAGTGACCA AAGGTGTGC   2089